Remote Sensing of Environment 308 (2024) 114190

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Learning spectral-indices-fused deep models for time-series land use and land cover mapping in cloud-prone areas: The case of Pearl River Delta

Zhiwei Li [a,b], Qihao Weng [a,b,c,*], Yuhan Zhou [a,b], Peng Dou [d], Xiaoli Ding [a,c]

[a] *Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong*
[b] *Research Centre for Artificial Intelligence in Geomatics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong*
[c] *Research Institute for Land and Space, The Hong Kong Polytechnic University, Hung Hom, Hong Kong*
[d] *Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou, China*

ARTICLE INFO

Editor: Marie Weiss

Keywords:
Land use and land cover
Time series images
Deep learning
Sentinel-2
Cloudy and rainy areas

ABSTRACT

Mapping of highly dynamic changes in land use and land cover (LULC) can be hindered by various cloudy conditions with optical satellite images. These conditions result in discontinuities in high-temporal-density LULC mapping. In this paper, we developed an integrated time series mapping method to enhance the LULC mapping accuracy and frequency in cloud-prone areas by incorporating spectral-indices-fused deep models and time series reconstruction techniques. The proposed method first reconstructed cloud-contaminated pixels through time series filtering, during which the cloud masks initialized by a deep model were refined and updated during the reconstruction process. Then, the reconstructed time series images were fed into a spectral-indices-fused deep model trained on samples collected worldwide for classification. Finally, post-classification processing, including spatio-temporal majority filtering and time series refinement considering land–water interactions, was conducted to enhance the LULC mapping accuracy and consistency. We applied the proposed method to the cloud- and rain-prone Pearl River Delta (i.e., Guangdong–Hong Kong–Macao Greater Bay Area, GBA) and used time series Sentinel-2 images as the experimental data. The proposed method enabled seamless LULC mapping at a temporal frequency of 2–5 days, and the production of 10 m resolution annual LULC products in the GBA. The assessment yielded a mean overall accuracy of 87.01% for annual mapping in the four consecutive years of 2019–2022 and outperformed existing mainstream LULC products, including ESA WorldCover (83.98%), Esri Land Cover (85.26%), and Google Dynamic World (85.06%). Our assessment also reveals significant variations in LULC mapping accuracies with different cloud masks, thus underscoring their critical role in time series LULC mapping. The proposed method has the potential to generate seamless and near real-time maps for other regions in the world by using deep models trained on datasets collected globally. This method can provide high-quality LULC data sets at different time intervals for various land and water dynamics in cloud- and rain-prone regions. Notwithstanding the difficulties of obtaining high-quality LULC maps in cloud-prone areas, this paper provides a novel approach for the mapping of LULC dynamics and the provision of reliable annual LULC products.

1. Introduction

Land use and land cover (LULC) datasets play a vital role as fundamental data in various applications, including land use planning and management, eco-environment conservation, and agriculture. LULC mapping has consistently been a popular research topic, and it continues to evolve alongside the advancements in data acquisition and processing capacities (Gómez et al., 2016; Ma et al., 2017; Talukdar et al., 2020; Vali et al., 2020). Over the past few decades, the spatial resolution of LULC mapping has been continuously improved from medium to high resolution at the meter- and even submeter-levels (Fan et al., 2020; Zanaga et al., 2021; Tong et al., 2023). Meanwhile, the temporal frequency of LULC mapping is also promoted from annual mapping to near real-time mapping (Gong et al., 2019; Zhang et al., 2021; Brown et al., 2022). Recent machine learning techniques, especially deep learning, have significantly revolutionized LULC mapping, and are widely used for producing new regional and global LULC products (Zanaga et al., 2021; Brown et al., 2022; Li et al., 2023). The advancements of LULC mapping in the above-mentioned aspects have marked a significant milestone in achieving accurate and continuous LULC mapping with

[*] Corresponding author at: Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong.
E-mail address: qihao.weng@polyu.edu.hk (Q. Weng).

https://doi.org/10.1016/j.rse.2024.114190
Received 27 October 2023; Received in revised form 26 April 2024; Accepted 30 April 2024
Available online 4 May 2024

dense image time series.

Image classification serves as the foundation of LULC mapping and categorizes remotely sensed images on the basis of various spatial units from pixels, objects, and geo-parcels to the scenes (Blaschke, 2010; Ma et al., 2017; Yang et al., 2017; Zhang et al., 2018). Over the previous decades, LULC classification methods that utilize spectral and spatial information from different sources of satellite remote sensing data, including optical and radar images, have been intensively developed (Foody, 1995; Yan et al., 2006; Myint et al., 2011; Duro et al., 2012; Qi et al., 2012). Traditional classification methods suffer from identifying complex land patterns, especially in urban areas, due to the limited number of low-level features in the spectral and spatial domains involved. Deep learning, emerging as a new focal point in machine learning, has brought about significant breakthroughs that have sparked a revolution in the remote sensing field (Zhang et al., 2016; Zhu et al., 2017; Yuan et al., 2020). Likewise, deep learning has facilitated LULC mapping and boosted its accuracy to state-of-the-art levels. The significant advantages of deep learning for image classification over traditional rule-based and machine learning methods are their strong feature representation ability, which allows them to learn from data and adaptively extract a huge number of discriminating features for classification (Zeiler and Fergus, 2014; Kussul et al., 2017; Li et al., 2018). The superiority and potential of deep learning that leverages diverse multi-scale and multi-level features extracted from images become ever more explicit and significant with the increase in the spatial resolution of satellite images for LULC mapping. Meanwhile, the integration of multi-source or multi-modal data has improved the LULC mapping performance. For example, the combined use of optical, SAR, and topography data can achieve higher LULC mapping accuracy than using mono-modal data alone (Hong et al., 2020; Li et al., 2022b). Herein, one of the advantages of deep learning is that it makes the fusion of heterogeneous multi-source and multi-modal data for LULC mapping easier than traditional classification methods.

Recently, deep learning-based LULC classification methods have been mostly used for pixel-to-pixel LULC classification (Li et al., 2016; Kussul et al., 2017; Scott et al., 2017; Ienco et al., 2019; Dou et al., 2021), which typically involve training end-to-end deep models that classify each pixel or segmented object of the input image into a specific class. Specifically, a number of spatial aggregation and boundary refinement strategies, including conditional random field (Fu et al., 2017), object-based image analysis (Zhang et al., 2018; Liu et al., 2019), skeleton decomposition (Huang et al., 2018), and hierarchical segmentation (Tong et al., 2020), are integrated with deep learning models to refine the classification results and preserve the completeness of the ground objects in the classification results. Furthermore, the geotagged photographs can be used as auxiliary data to facilitate the LULC classification and validation (Tracewski et al., 2017; Xu et al., 2017; Xing et al., 2018). In particular, dual-branch convolutional neural network (CNN) architectures have been designed in several studies to better deal with panchromatic and multi-spectral bands separately (Gaetano et al., 2018; Huang et al., 2018). More recently, the benefits of combining two different networks, namely, CNN and recurrent neural network (Interdonato et al., 2019; Qiu et al., 2019a) and CNN and Transformer (Wang et al., 2022; Song et al., 2023), have also been exploited to enhance the feature representation and improve the model performance for LULC classification. Recent public LULC products, including Esri Land Cover (Karra et al., 2021), Google Dynamic World (Brown et al., 2022), and SinoLC-1 (Li et al., 2023), have been developed using deep learning approaches. These advancements highlight the growing popularity of deep learning methods in practical applications, particularly demonstrating their effectiveness in challenging classification scenarios, such as for cropland and grass/shrub classes (Wang and Mountrakis, 2023).

Despite the remarkable progress made in recent years, two major issues in the field of LULC mapping persist. On one hand, the cloud coverage in optical image time series reduces the availability of data for time series LULC mapping. Meanwhile, the importance of accurate cloud masks cannot be overstated for precise high-temporal-density near real-time LULC mapping, especially in cloud-prone areas. However, the cloud masks that are commonly used are often not highly accurate, as evidenced by multiple recent studies (Baetens et al., 2019; Sanchez et al., 2020; Tarrio et al., 2020), leaving space for further improvements. Specifically, for Sentinel-2 imagery, the Sentinel-2 Level 1-C cloud mask product has been found to generally underestimate cloud presence, which cannot be ignored (Coluzzi et al., 2018). Other existing cloud detection methods, such as Sen2Cor (Richter et al., 2012), MAJA (Hagolle et al., 2017), and Fmask (Qiu et al., 2019a, 2019b), exhibit varying limitations in accurately distinguishing clouds from bright ground surfaces and in effectively identifying thin cirrus clouds and cloud shadows, as summarized in the study by (Tarrio et al., 2020). These limitations highlight the necessity for more sophisticated methods, such as the deep learning model by Li et al. (2021), which offers improved accuracy through the use of multiscale features but is limited by the requirement for extensive training data. Despite the progress made, the accuracy of cloud masks remains suboptimal, underscoring the necessity for ongoing efforts to refine cloud detection techniques to enhance LULC mapping capabilities. Additionally, revealing the quantitative effects of clouds and different cloud masks on the accuracy of LULC mapping is another aspect that warrants further exploration (Tarrio et al., 2020; Ling et al., 2021). On the other hand, the identification of dynamically changing land patterns, especially over varying water areas, such as paddy fields, is challenging but important for the composition of accurate annual LULC maps (Waleed et al., 2022). Meanwhile, the annual LULC mapping, which leverages all available image time series within a year, is expected to further promote the mapping accuracy.

To improve high-temporal-density LULC mapping in cloudy and rainy areas, we proposed an integrated time series LULC mapping method to enhance the LULC mapping under dense cloud coverage and varying water conditions. This method aims to generate and composite seamless near real-time, monthly, seasonal, and annual LULC maps with high accuracy. Specifically, spectral-indices-fused deep models that fuse task-specific spectral indices from images are constructed for cloud masking and LULC classification, respectively. The refined cloud masks through time series refinement are expected to reduce the negative influences of clouds on LULC mapping. Meanwhile, the reconstruction of time series cloudy images will benefit LULC mapping in terms of accuracy. In particular, the consideration of temporal change patterns in post-classification processing benefits the identification of classes, such as crops, which may frequently occur in water–land interactions. The objectives of this study are as follows: 1) develop an integrated method for high-quality time series LULC mapping in rainy and cloudy areas; 2) reveal the effects of clouds on LULC mapping and the benefits of time series reconstruction and LULC mapping with dense image time series; and 3) produce a series of LULC products over the study area that outperform the other existing products. The proposed method is expected to be applied in other regions in the world to generate highly reliable LULC products, especially in cloud-prone areas.

2. Study area and data

The Guangdong–Hong Kong–Macao Greater Bay Area (GBA thereafter) (Fig. 1), one of the most developed regions in China, encompasses a total of 11 major cities and spans in the Pearl River Delta region, which has experienced rapid land use and land cover changes in the recent decades (Weng, 2002; Zhang and Weng, 2016; Zhang et al., 2017). The rainy and cloudy weather conditions in the GBA present a challenge for high-temporal-density LULC mapping, especially during annual rainy seasons when the region experiences dense cloud coverage. The existence of the Pearl River and coastal environment results in frequent land and water interactions in the GBA, leading to highly dynamic changes in LULC types in the region. Thus, GBA is selected as the study area to examine the effectiveness of the proposed method for LULC mapping in

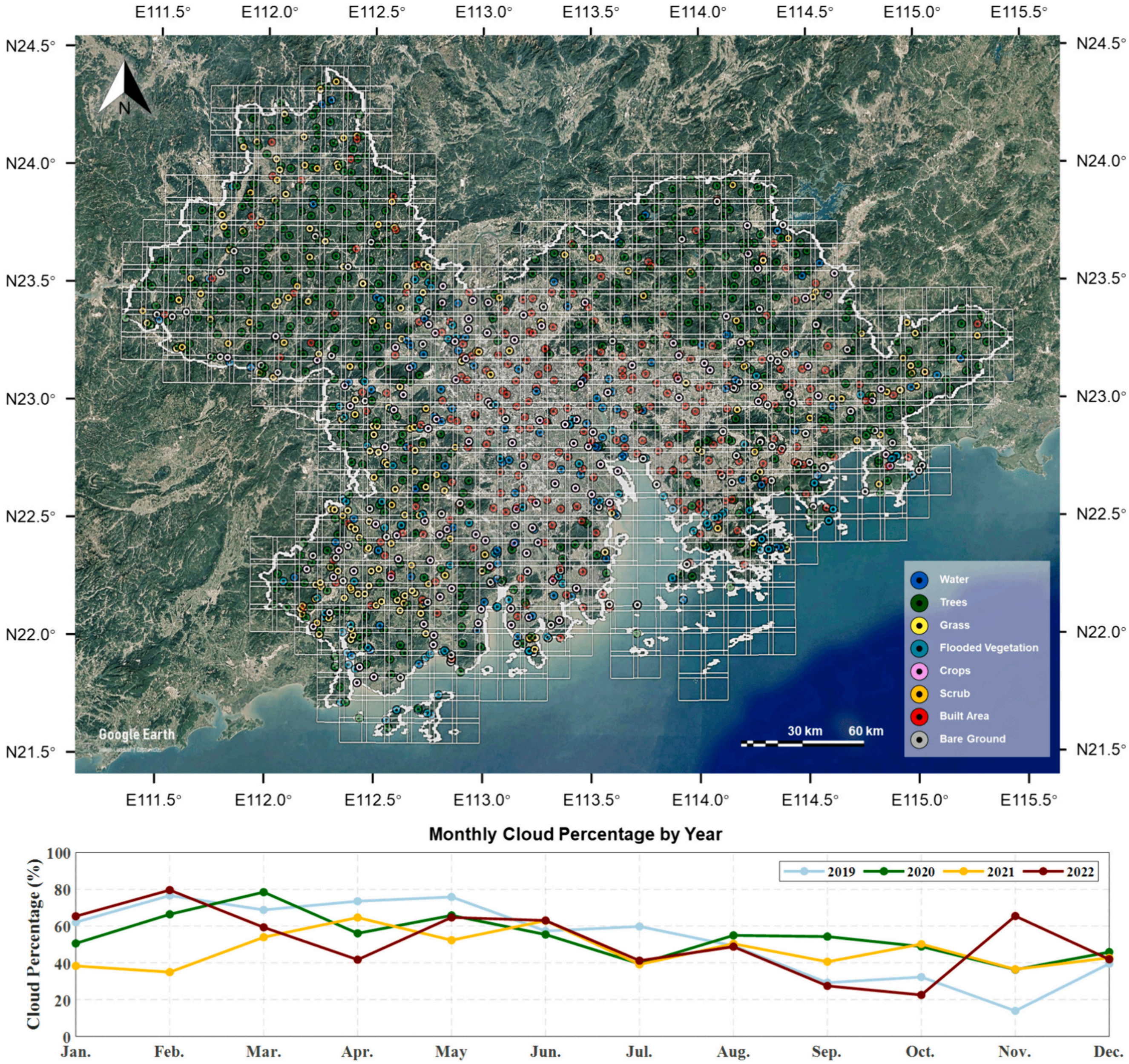

Fig. 1. Study area Pearl River Delta and monthly cloud percentage in Sentinel-2 imagery utilized for the study. The upper image shows the location of the study area and the distribution of validation sample sites. The lower image represents the average monthly percentages of cloud coverage in Sentinel-2 imagery in the study area during 2019–2022.

cloud-prone areas.

The Sentinel-2 images from the Sentinel-2 A/B constellation operated by the European Space Agency are selected as the study data. Specifically, the harmonized Sentinel-2 Level-2 A (i.e., surface reflectance) image time series in the GBA is exported through the Google Earth Engine (GEE) platform (Gorelick et al., 2017). In addition, the LULC mapping experiments incorporated an image dataset spanning 4 years from December 1, 2018, to January 31, 2023. The spatial resolution of all exported spectral bands in Sentinel-2 images is unified to 10 m by the default nearest neighbor resampling in GEE, and the temporal resolution is 2–5 days, varying from different regions. Majority of the areas in the GBA are revisited by Sentinel-2A/B satellites every 5 days, resulting in a total of 299 typical coverages during the study period. The Sentinel-2 time series over GBA among the 4 years are densely covered by clouds with an estimated mean cloud percentage as high as 51.61%. The monthly mean cloud percentages for each year are provided in Fig. 1, which depicts that high-quality and high-temporal-density LULC mapping in GBA is challenging due to the cloud coverage. To efficiently proceed with dense time series image data, as shown in Fig. 1, the Sentinel-2 time series in the GBA is divided into 40 × 30 image tiles for tile-by-tile data acquisition and processing. Only the 695 tiles that cover the GBA are involved for experiments, each tile with a size of approximately 1360 × 1278. Each of the two neighboring image tiles overlaps by 0.01° in all four directions to alleviate the artifacts at the edges of results produced by deep models and avoid inconsistency at image edges when mosaicing the image tiles to an entire large image for GBA.

We collected sample points based on very high-resolution satellite images in Google Earth through manual interpretation to comprehensively evaluate the time series LULC products. In Fig. 1, the 1263 sample sites over GBA are selected by stratified random sampling. The number and percentages of sample sites for each majority class are as follows: water (116, 9.18%), trees (591, 46.79%), grass (41, 3.25%), flooded vegetation (60, 4.75%), crops (117, 9.26%), scrub (47, 3.27%), built area (245, 19.40%), and bare ground (46, 3.64%). The LULC labels of the sample points are double-checked by two individual experts to guarantee the labeling quality. Meanwhile, we positioned the selected sample points of each site at the center of the homogenous area to alleviate the labeling errors caused by the spatial misalignment and spatial resolution differences between Sentinel-2 and Google Earth images. The selected sample points cover multiple dates between December 2018 and January 2023, with consideration for the LULC-changed areas. Accordingly, multiple labels are associated with different dates for sample points where LULC changes occurred during the above-mentioned period. Meanwhile, only a single label is associated without a specific date for sample points that belong to the same LULC category over the entire study period, if all manually interpreted labels are in the same category based on all available historical observations on Google Earth during the study period. Among the 1263 sample sites, 1188 are LULC-unchanged sites, and 75 sites have undergone LULC changes, from which 1188 and 186 LULC labels are collected, respectively. Such a collection of sample points, including the LULC-changed and unchanged areas over time, will guarantee a more comprehensive evaluation of time series LULC products.

3. Methodology

We proposed an integrated LULC mapping method, which comprises four main steps (Fig. 2). The proposed method first initializes cloud and cloud shadow masks for time series Sentinel-2 images by the spectral-indices-fused deep model based on CNN. The cloud- or cloud shadow-contaminated pixels in the Sentinel-2 time series are then reconstructed through time series filtering, during which the initial cloud masks are refined to improve the reconstruction effects. Thereafter, the reconstructed time series images are fed into another deep model trained on samples collected worldwide for LULC classification. Finally, post-classification processing is conducted to enhance the LULC mapping accuracy and consistency. The accuracy of the produced LULC products is quantitatively evaluated and compared with existing mainstream

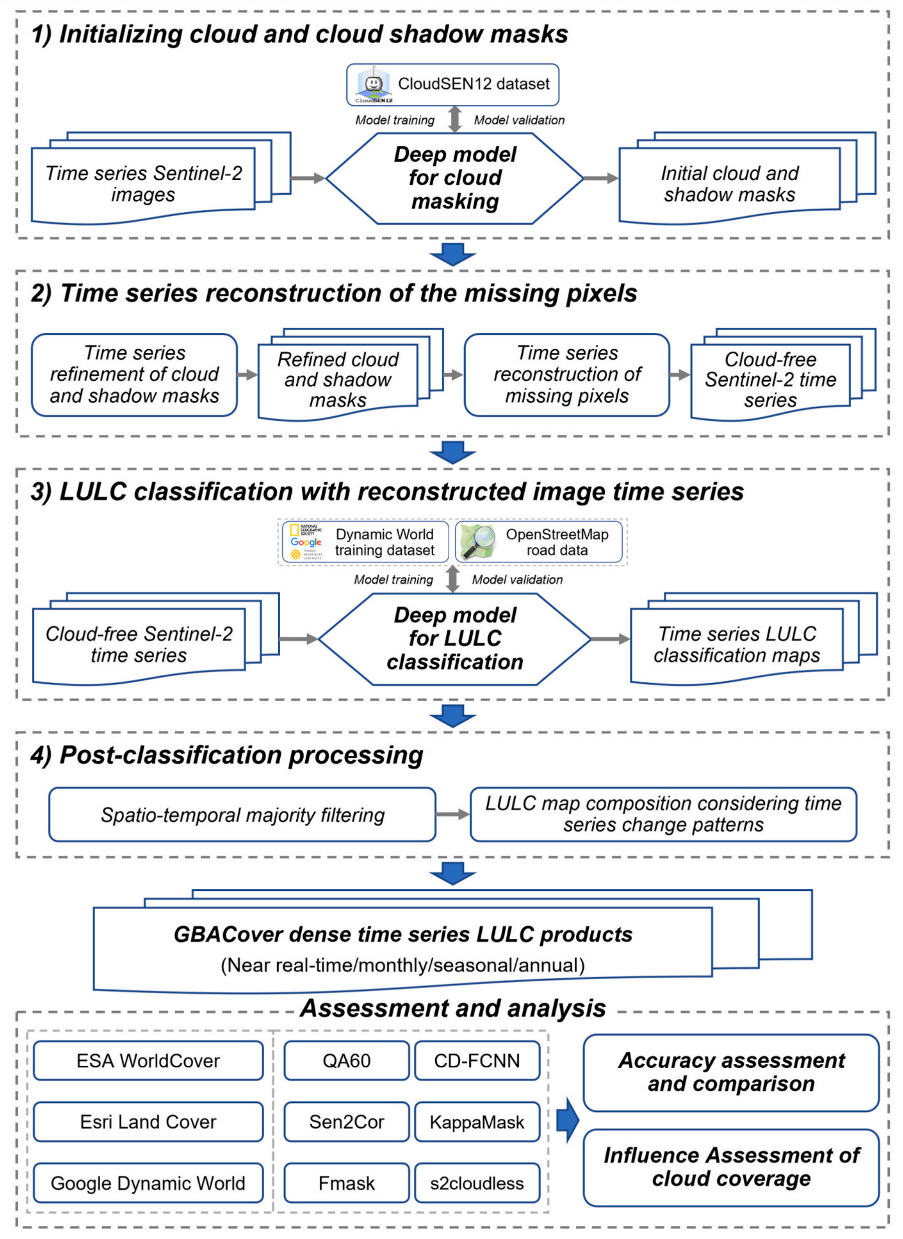
1) Initializing cloud and cloud shadow masks
CloudSEN12 dataset
Model training
Model validation
Time series Sentinel-2 images
Deep model for cloud masking
Initial cloud and shadow masks
2) Time series reconstruction of the missing pixels
Time series refinement of cloud and shadow masks
Refined cloud and shadow masks
Time series reconstruction of missing pixels
Cloud-free Sentinel-2 time series
3) LULC classification with reconstructed image time series
Dynamic World training dataset
OpenStreetMap road data
Model training
Model validation
Cloud-free Sentinel-2 time series
Deep model for LULC classification
Time series LULC classification maps
4) Post-classification processing
Spatio-temporal majority filtering
LULC map composition considering time series change patterns
GBACover dense time series LULC products
(Near real-time/monthly/seasonal/annual)
Assessment and analysis
ESA WorldCover
Esri Land Cover
Google Dynamic World
QA60
CD-FCNN
Sen2Cor
KappaMask
Fmask
s2cloudless
Accuracy assessment and comparison
Influence Assessment of cloud coverage

Fig. 2. The flowchart of the proposed time series LULC mapping method.

LULC products in the GBA. In particular, the effects of the cloud coverage on LULC mapping are evaluated.

3.1. Initializing cloud and cloud shadow masks

The cloud masks provided in the quality assessment bands of Sentinel-2 images are rough and do not label the cloud shadow areas, which will lead to significant biases for the following image reconstruction and classification steps. Accordingly, we used a dual-branch deep model for cloud and cloud shadow detection in Sentinel-2 Level-2 A images to produce high-quality masks to support the precise reconstruction of missing areas in time series images.

With regard to the cloud and cloud shadow detection task, we used a large Sentinel-2 cloud and cloud shadow detection dataset, i.e., CloudSEN12 (Aybar et al., 2022), for the model training, in which the samples collected worldwide and with varying cloud coverage conditions will guarantee the state-of-the-art model performance and model generation ability across different image scenarios. In the CloudSEN12 datasets, 10,000 Sentinel-2 image patches on 2000 regions of interest (ROIs) worldwide are labeled with high confidence. There are five Sentinel-2 image patches in each ROI, each with a size of 509 × 509. We included four of the five samples in each ROI for model training. The remaining one sample, which is randomly selected from the five samples in each ROI, is used for the validation of the model performance. The Sentinel-2 images in the CloudSEN12 dataset are collected at both Level-1C and Level-2 A, and the label covers four classes, including thick cloud, thin cloud, cloud shadow, and clear sky. In this paper, the model trained with Sentinel-2 Level-2 A images and corresponding labels can distinguish cloud and cloud shadow from clear sky in the Sentinel-2 surface reflectance data for experiments.

In Fig. 3, the designed dual-branch Spectral-Indices-Fused Deep Model (SIFDM) is a U-Net-like (Ronneberger et al., 2015) encoder–decoder architecture and consists of dual-branch encoders, which are used to extract multi-level features from the original images and their derived spectral indices that contain the prior knowledge, and a decoder, which is designed to step-wisely fuse the multi-level features and output the desired cloud and cloud masks. The underlying motivation for this design is to leverage the knowledge inherent in the original images by explicitly providing this information, thereby guiding the training process and amplifying the model's ability to discern the intricate patterns of clouds and their shadows. The derived spectral indices from the original images, including the haze optimized transform (HOT) index (Zhang et al., 2002; Zhu and Woodcock, 2012), visible band ratio (VBR) (Li et al., 2017), cloud displacement index (CDI) (Frantz et al., 2018), and cloud shadow index (Zhai et al., 2018), are discriminative features for cloud and cloud shadow masking and have been widely used in previous studies.

Specifically, five levels of features are extracted and fused in the model. The detailed network structure of the designed SIFDM model is provided in Appendix Table 1. In the encoder module, each level of features in the encoder is generated by double convolutions with a 3 × 3 convolution kernel size. A batch normalization layer and a rectified linear unit (ReLU) activation function are followed after each convolution layer. Meanwhile, a pooling layer is added after the first four double convolution blocks to reduce the spatial dimensions of the feature through a maximum operation over a 2 × 2 window. When the spatial dimensions of the feature maps are reduced by half from the second to the last block in the encoder, the numbers of feature maps at the five levels are set to {64, 128, 256, 512, 512} to retain more information as needed. In the decoder module, features at different levels are step-wisely fused and recovered to the same height and width as the input image. The same-level features from each branch of the encoder are concatenated and upsampled to twice the height and width of the input features through bilinear interpolation. Thereafter, the unsampled features are fed into the same double-convolution block as in the encoder. The number of output feature maps will be reduced to 1/4 or 1/2 adaptively after the double convolutions. The output feature maps are concatenated with the features from each branch of the encoder at the next level and then fed into a double-convolution block after being upsampled. The above-mentioned process is repeated until the last output, which is generated by a single convolution layer with a size of 1 × 1 kernel, to retain more spatial details in the final output maps. The numbers of output feature maps of each double-convolution block, excepting the last convolution layer in the decoder, are set to {256, 128, 64, 64}. Meanwhile, the number of output features for the last convolution layer is determined by the specific task at hand and is based on the number of classes that are labeled in the training datasets. Here, a cross-entropy loss function is employed to supervise the model training. Overall, such a design leverages the derived spectral indices that contain the prior knowledge, enabling the model to accurately discern complex patterns of clouds and their shadows, thereby facilitating precise image reconstruction in the subsequent steps.

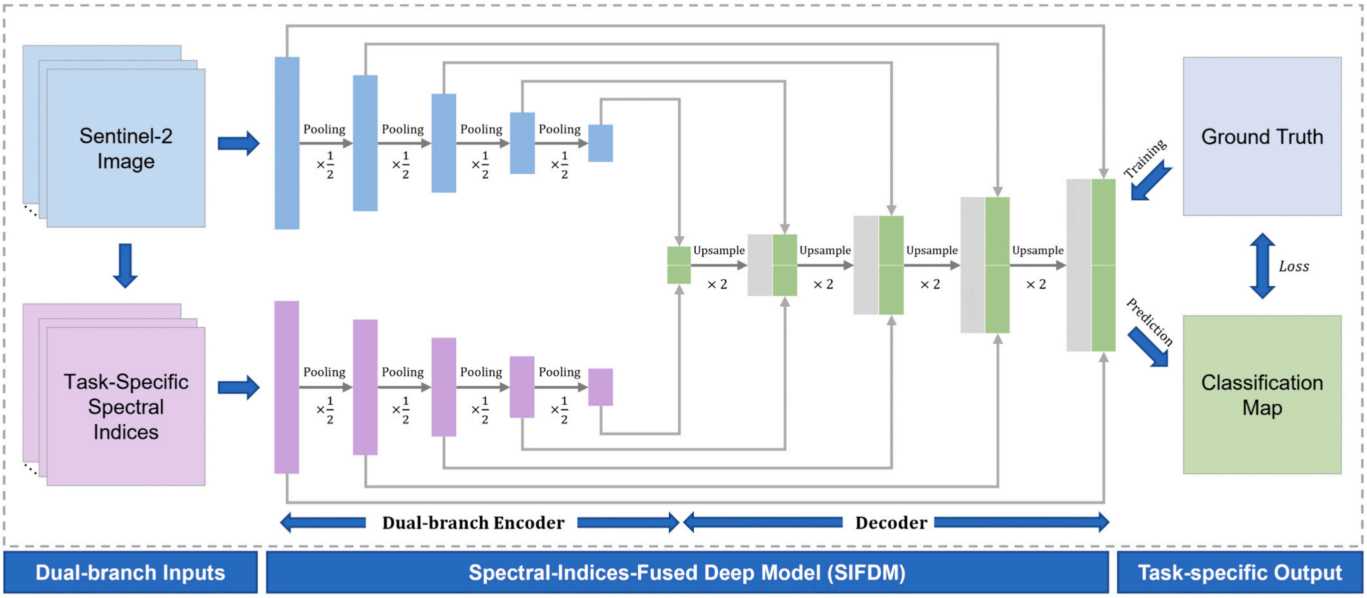

Fig. 3. The network architecture of the dual-branch spectral-indices-fused deep model.

During the model training, a stochastic gradient descent (SGD) optimizer and a poly learning rate decay policy were chosen. In this setup, the learning rate follows a polynomial decay from an initial value of 0.1 to zero when reaching the maximum iteration limit. Finally, the model was adequately trained for 100,000 iterations with a batch size of 8 on the 8000 training samples. In the application stage of the pre-trained cloud and cloud shadow masking model, we initialize a binary mask $M_{initial}$ based on the model output, in which the invalid cloud/shadow-contaminated pixels are denoted as one, while the valid clear-sky pixels are labeled as zero.

3.2. Time series reconstruction of the missing pixels

To reconstruct the missing pixels contaminated by clouds and cloud shadows, we utilized a Whittaker filter (Whittaker, 1922; Eilers, 2003) to conduct time series filtering for each pixel band by band. Whittaker filtering has been widely applied for time series signal smoothing in various fields due to its superior smoothing capability and high computation efficiency. Given the reflectance of pixel time series $\boldsymbol{y}$ with a length of n (i.e., the total number of pixels), the goal of Whittaker filtering is to obtain a smoothed time series z with high fidelity. Thus, the objective function in Eq. (1) can be constructed to find the time series z that minimized Q.

$$Q = |\boldsymbol{y} - z|_2^2 + \lambda |\boldsymbol{D}z|_2^2 \tag{1}$$

where the first fidelity term $|\boldsymbol{y} - z|_2^2$ measures the usual sum of squares of differences between $\boldsymbol{y}$ and z, the second smoothing term $|\boldsymbol{D}z|_2^2$ can be expressed with the sum of squares of second-order differences, and λ is the weight parameter used to balance the fidelity term and smoothing term. Thus, the above-mentioned object function can also be written as Eq. (2).

$$Q = \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{y}_i - z_i)^2 + \lambda\sum_{i=2}^{n-1}(z_{i-1} - 2z_i + z_{i+1})^2 \tag{2}$$

Before inputting the reflectance of the pixel time series $\boldsymbol{y}$ over the entire study period for filtering, the missing values in $\boldsymbol{y}$ are pre-filled through linear interpolation. This process estimates the missing values by interpolating along the line segment between two adjacent known data points. Here, only the spectral bands used for the subsequent LULC classification are engaged in reconstruction. Although the generated masks by the deep model are highly accurate, they are not ideal, and omission errors, such as the thin clouds in images, may still occur, which will lead to biases in the time series reconstruction step. To this end, time series filtering is utilized to leverage temporal information to refine masks. Specifically, with regard to the reflectance of each pixel time series, the invalid pixels in the time series labeled in the initial mask $M_{initial}$ are first filled through the nearest neighbor interpolation. Thereafter, triple upper-enveloped Whittaker filtering is conducted on the interpolated time series, as shown in Fig. 4, in which the values of the pixels in the time series increased after filtering will be set to the original value before filtering in the first two rounds of filtering. This upper-enveloped strategy specifically targets the omission of clouds, because clouds both occupy larger portions of cloud-covered images and generally lead to more significant changes in pixel reflectance than cloud shadows. The purpose of such triple filtering is to obtain the reference clear-sky reflectance curve of the target pixel. The reflectance curve is relatively smooth, and noises are largely filtered out. Finally, if the absolute value difference for each pixel in the time series before and after triple filtering is greater than a threshold that is empirically set to 0.04, and the pixel value before filtering is larger than the 80% quantile value or smaller than the 20% quantile values of all valid pixels in the time series indicated in $M_{initial}$, such a pixel in the time series will be labeled as the invalid cloud or cloud shadow pixel and is merged into the initial mask $M_{initial}$ to obtain the refined mask $M_{refined}$.

The refined masks $M_{refined}$ will be used as a reference to guide the reconstruction of the missing pixels in the time series. Specifically, the invalid values in the pixel time series are also first filled by nearest neighbor interpolation. Afterward, Whittaker filtering is applied to smoothen the time series. Although the refined masks are not completely accurate, the omission error has been significantly reduced (Fig. 4). Such

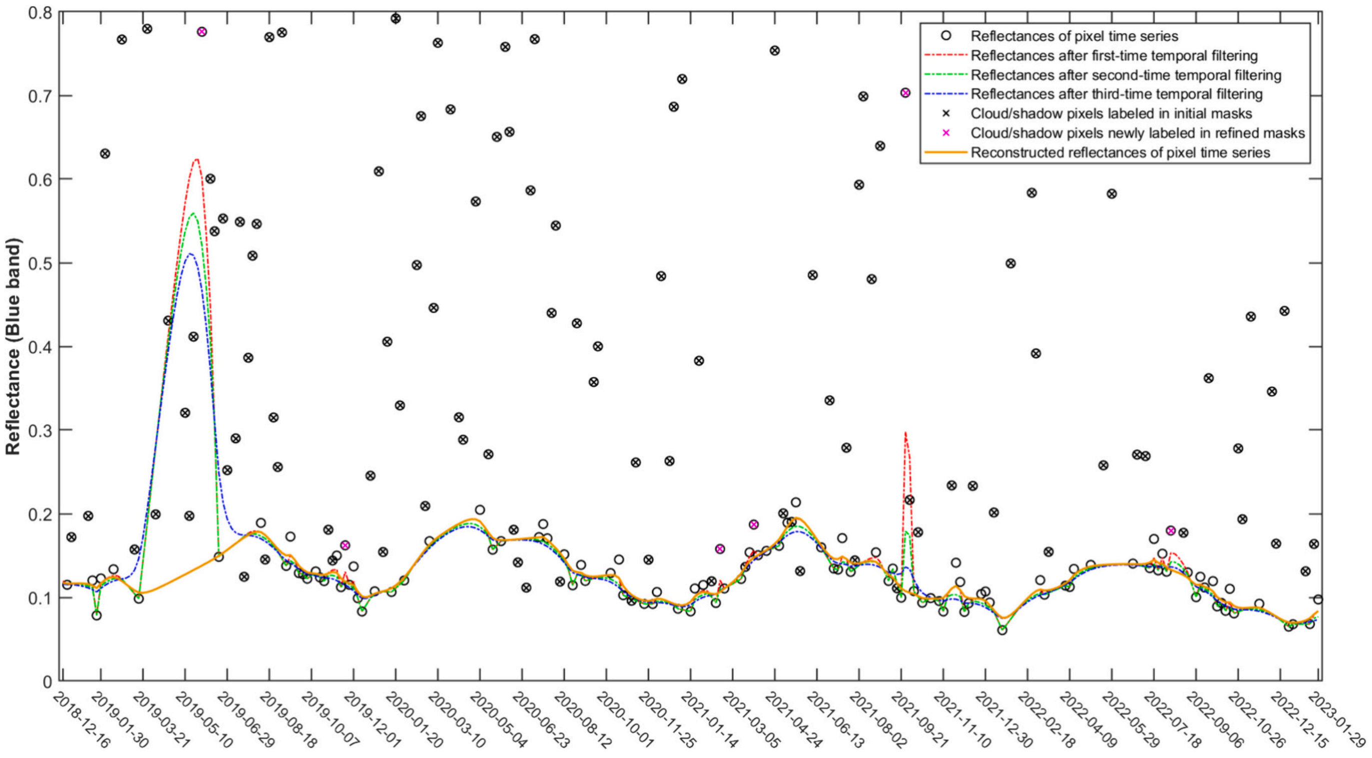

Fig. 4. Time series refinement of the cloud and cloud shadow masks and reconstruction of the missing pixels.

a mask refinement will ensure more satisfactory reconstruction results, although it may sacrifice a few valid pixels that are mistakenly labeled as cloud or cloud shadow. This phenomenon occurs because omission errors in the cloud and cloud shadow masks are more serious than commission errors for time series fitting (Zhu and Woodcock, 2012, 2014). The weight parameter λ is set to four for enhanced smoothing effects in identifying omitted cloud or shadow pixels during mask refinement, and to two to guarantee high fidelity of pixel reflectance before and after time series filtering. The reconstructed cloud-free image time series will be used for the subsequent LULC classification.

3.3. LULC classification with the reconstructed image time series

This paper adapted the same network architecture (Fig. 2) for LULC classification, which has been used for cloud masking. This initiative is carried out because the LULC classification and cloud masking are semantic segmentation tasks. The differences between the models for the above-mentioned two tasks are the inputs and outputs. The inputs for the LULC classification model at the dual-branch encoder are spectral bands of reconstructed cloud-free Sentinel-2 image and derived spectral indices. Specifically, the input spectral bands include B2, B3, B4, B5, B6, B7, B8, B8A, B11, and B12 (i.e., all spectral bands except aerosol band B1 and water vapor band B9). The input-derived spectral indices for classification, which refer to the Sentinel-2 Level-2 A product algorithm (Richter et al., 2012), include normalized difference vegetation index (NDVI), normalized difference water index (NDWI), normalized difference built index (NDBI), normalized difference snow index (NDSI), and ratios of band 2/band 4, band 8/band 3, band 2/band 11, and band 8/band 11. The derived spectral indices are utilized to diversify and enrich the input features and enhance the model performance. The output of the LULC classification model is the multi-class classification map. The loss function used to supervise the model training and measure the difference between the model outputs and the ground truths is the linear combination of focal loss (Lin et al., 2020) and dice loss (Milletari et al., 2016), as introduced in previous studies (Cheng et al., 2021; Kirillov et al., 2023). Although the focal loss is helpful in boosting the model performance on the hard-classified samples, the dice loss benefits the model performance on the classes with few percentages of samples. Such a loss combination will contribute to the construction of well-balanced losses for model training supervision and better model performance.

The Dynamic World training dataset for global LULC categorization (Tait et al., 2021), which is constructed based on over 5 billion pixels of manually labeled Sentinel-2 images collected at global sites, is used for LULC classification model training and validation. The dataset was created under the *National Geographic Society - Google - World Resources Institute Dynamic World* project. The 10 m resolution image patches of the datasets with a size of 510 × 510 pixels are densely labeled using a 10-category classification schema. Majority of these image patches were acquired in 2019, with the exception of approximately 10% of the patches obtained in 2017, primarily from cloud-prone areas worldwide. Note that cloud is marked in labels for the image patches contaminated by clouds. With regard to the proposed classification model in this study, a total of 24,528 Sentinel-2 image patches and their corresponding pixel-level labels from both expert and non-expert labeling sources were involved for model training. The total number of valid training pixels is approximately 5.73 billion, with the following pixel percentages for each class: water (8.63%), trees (38.99%), grass (2.42%), flooded vegetation (2.43%), crops (13.12%), scrub (34.05%), built area (4.74%), bare ground (3.43%), and snow/ice (3.01%). Additionally, an extra 409 samples with the same image size are used for model validation. The total number of valid validation sample pixels is approximately 29.46 million and the pixel number percentages for each class are as follows: water (22.34%), trees (29.40%), grass (1.30%), flooded vegetation (1.39%), crops (25.37%), scrub (5.47%), built area (9.18%), bare ground (3.49%), and snow/ice (2.06%). The highway data in OpenStreetMap (OSM) datasets (OSM, 2017) are rasterized with 1 pixel dilatation and merged into the class of built area in training dataset to enhance the ability in identifying roads in built areas.

During the LULC classification model training stage, the same strategy for training the deep model for cloud and cloud shadow detection, as outlined in Section 3.1, is followed with only two differences: a reduced batch size of 4 and an increased maximum number of iterations to 200,000. Once the classification model is trained, the large-size image can be processed with the pre-trained model patch by patch.

3.4. Post-classification processing

The post-classification processing is conducted to refine the classification results and improve the consistency of the time-series LULC maps. Efforts in two aspects are made to achieve this goal.

3.4.1. Improving the LULC mapping consistency through spatio-temporal majority filtering

A 3D majority filter with a kernel size of 3 × 3 × 3 is employed to filter out the spatio-temporal noises from the time-series LULC maps. The classes that most frequently occur within the 3 × 3 × 3 sliding window will be assigned as the class of the center pixel within the window. Such an operation is helpful in reducing the classification errors brought by random noise in the original images and inaccurate predictions by deep models. The time series consistency of LULC maps will be improved by using the spatio-temporal majority filtering.

3.4.2. LULC map composition considering time series change patterns

Temporal composition is conducted to generate monthly, seasonal, and annual LULC maps, in which the time series LULC maps at different lengths of time periods (i.e., one month, one season, and one year) are composited. The class value for each pixel most frequently occurring in the time series of different lengths will be considered the final LULC category in the monthly, seasonal, or annual LULC map. In particular, the temporal change patterns are considered during the LULC map compositions, especially in areas where land and water interactions frequently occur. Specifically, the LULC category for a pixel LULC time series changes between crops and water more than two times within one year. The final category of such pixel will be assigned as crops in the annually composited map. Such an LULC map refinement rule is also applicable for composting seasonal maps. Based on the near real-time time series LULC maps and the above-mentioned post-processing, the monthly, seasonal, and annual LULC maps can be generated. Thus, the LULC dynamics can be monitored at different frequencies.

4. Results and analysis

The 4-year Sentinel-2 time series covering GBA was used as the experiment data, in which a total of 695 tiles of image time series exist. All image time series tiles are processed tile by tile according to the processes shown in Fig. 1 by using the pretrained cloud masking model and LULC classification model. The monthly, seasonal, and annual LULC maps for the entire GBA can be finally produced by composting and mosaicking time-series LULC map tiles. In this section, we will first evaluate the performance of the deep models, conduct an accuracy assessment of our produced LULC products, and compare them with the mainstream public LULC products. Finally, the effects of the cloud coverage on LULC mapping will be investigated. The five metrics, including overall accuracy (OA), producer's accuracy (PA), user's accuracy (UA), mean intersect over union (mIoU), and F-score, are involved in the accuracy assessment.

4.1. Accuracy validation of deep models for cloud masking and LULC classification

4.1.1. Accuracy validation of the cloud and cloud shadow masking model

Based on the 2000 randomly selected global validation samples from the 2000 ROIs worldwide in the CloudSEN12 dataset (Aybar et al., 2022), the accuracy of the trained cloud and cloud shadow masking model can be quantitatively assessed and compared with other methods. The compared methods include, QA60 (i.e., quality assessment bands associated with Sentinel-2 images) (Chambrelan, 2012), Sen2Cor (Richter et al., 2012), s2cloudless (Zupanc, 2017), Fmask (Qiu et al., 2019a, 2019b), CD-FCNN (López-Puigdollers et al., 2021), and KappaMask (Domnich et al., 2021).

Note that the initial cloud and cloud masks, obtained using the SIFDM model and without refinement, are utilized for accuracy evaluation. Among the compared deep learning models, CD-FCNN and KappaMask only use Sentinel-2 image bands as input, while SIFDM with a dual-branch encoder incorporates both image bands and their derived spectral indices as input. Additionally, all cloud and cloud shadow masks generated by the compared methods are already included in the CloudSEN12 dataset (Aybar et al., 2022) and thus directly collected for comparisons, without involving model training or fine-tuning. We used thresholds of 40 and 500 to binarize the grayscale cloud masks generated by s2cloudless and CD-FCNN, respectively. The optimal segmentation thresholds are selected by a rough parameter sensitivity analysis. The threshold that led to the highest overall accuracy will be used as the segmentation threshold to generate binary cloud masks for accuracy assessment. All subclasses of cloud in masks, including thick and thin clouds in the ground truth masks, opaque cloud and cirrus cloud in the QA60 masks, medium- and high-probability cloud, and thin cirrus in Sen2Cor, are merged into a single category of cloud. Meanwhile, the masks generated from QA60 or by s2cloudless and CD-FCNN only contain cloud information but without cloud shadow. To ensure fair comparisons of masks generated by different methods, we conduct accuracy assessments for cloud and cloud shadow separately to guarantee that the accuracies of cloud or cloud shadow in different masks are comparable. Specifically, the cloud and all other classes (including cloud shadow) will be treated as two classes for the accuracy assessment of cloud. The same approach is applied for the cloud shadow accuracy assessment.

Table 1 shows the details of the accuracy assessment results for the different methods. The proposed DKDFM model achieved the best performance in the cloud and cloud shadow detection on the 2000 test samples, followed by Fmask and KappaMask, which outperformed the other methods. The s2cloudless, CD-FCNN, KappaMask, and our SIFDM model are all machine-learning-based methods, and the last three methods are based on CNNs. Among all the compared methods, Fmask is the only non-machine-learning method, the cloud detection results of which are comparable with the machine learning methods that generally exhibited better performance in distinguishing clouds from bright non-cloud objects (Li et al., 2022a).

4.1.2. Accuracy validation of the LULC classification model

The additional 409 validation samples in the Dynamic World training dataset (Tait et al., 2021) with the same image size as training samples and collected from different biomes worldwide are used for model performance evaluation. Each image in the validation samples is labeled by three expert annotators and one non-expert annotator. Thereafter, the ground truths for the accuracy assessments can be composited based on the four individual annotations using the voting scheme. In this paper, ground truths composed using the voting scheme "Three Expert Strict", as detailed in the study by Brown et al. (2022), are used for model validation. This scheme stipulates that valid labels must be those where all three expert annotators labeled and all agreed. We validate our LULC classification model (i.e., SIFDM) and compare it with the model (named DW-Net hereafter) used to generate Dynamic World LULC products by using the same validation samples. The accuracy assessment results are listed in Table 2. The results show that the SIFDM model outperforms DW-Net in the overall LULC classification accuracy and achieves a higher overall accuracy of 93.39% than the 88.40% of DW-Net. The validation results also suggested that the classification accuracy of SIFDM is higher than that of DW-Net in terms of F-score for most classes, except for bare ground, indicating the potential of SIFDM for application in LULC mapping in GBA.

4.2. Accuracy assessment and comparison of the LULC products in GBA

The time-series LULC products in the GBA can be generated through the process chain introduced in the methodology section by using the pre-trained deep models. The LULC categories involved in GBA include water, trees, grass, flooded vegetation, crops, scrub, built area, and bare ground. We named the generated LULC products in GBA as GBACover, which include a series of dense time series LULC maps at multiple temporal densities (i.e., near real-time, monthly, seasonal, and annual scales). The accuracy of the generated GBACover products can be quantitatively evaluated and compared with other global LULC products over the GBA, including ESA WorldCover (Zanaga et al., 2021), Esri Land Cover (Karra et al., 2021), and Google Dynamic World (Brown et al., 2022) by using the manually interpreted LULC samples in GBA. All three existing LULC products were produced based on machine learning methods, making them highly accurate and comparable with LULC maps generated with the proposed method. The accuracies of the generated LULC maps are examined using the entire manual validation samples. The sample pixels that undergo changes over the study period are annotated with multiple labels at different dates. Meanwhile, the pixels without LULC changes are used to evaluate the accuracy of the entire time series LULC classification results. Specifically, the accuracy of the composited annual LULC maps can be assessed with validation samples within a specific year and compared with the existing products in quantitative assessment and visual inspection manners.

Two similar but different categories in the four different LULC products are involved in the comparison (i.e., flooded vegetation and wetlands). Detailed comparisons of class definitions among the LULC products are provided in Appendix Table 2. Esri Land Cover, Google Dynamic World, GBACover, and annotated samples contain flooded

Table 1
Accuracy comparison of the cloud and cloud masks generated by SIFDM and benchmarked methods.

Methods	Cloud				Cloud shadow			
	OA	PA	UA	mIoU	OA	PA	UA	mIoU
QA60	79.12% (+1.15%)	60.44% (−12.75%)	69.81% (+13.18%)	0.479 (−0.045)	\	\	\	\
Sen2Cor	84.55% (0.88%)	69.89% (−7.11%)	79.04% (+8.85%)	0.590 (−0.012)	90.48%	18.21%	84.70%	0.176
s2cloudless	89.02%	79.33%	85.10%	0.697	\	\	\	\
Fmask	88.82%	89.15%	78.55%	0.717	90.12%	46.61%	57.21%	0.346
CD-FCNN	88.57%	80.56%	82.96%	0.691	\	\	\	\
KappaMask	87.13% (−1.03%)	92.76% (−19.26%)	75.07% (+8.41%)	0.709 (−0.071)	91.50%	47.79%	66.78%	0.386
SIFDM (this paper)	**94.80%** (+0.38%)	**94.30%** (−3.89%)	**89.85%** (+4.36%)	**0.852** (+0.004)	**94.71%**	**62.55%**	**86.43%**	**0.570**

Note: The accuracy assessments were conducted for cloud and cloud shadow separately due to the unavailability of cloud shadow information in masks derived from QA60, s2cloudless, and CD-FCNN. The values in brackets represent the changes in accuracy attributed to the exclusion of thin clouds in QA60, Sen2Cor, KappaMask, and SIFDM.

Table 2
Accuracy comparison of the classification models for LULC mapping on the validation samples from Dynamic World training dataset.

Class (% of pixels)		Water (22.34%)	Trees (29.40%)	Grass (1.30%)	Flooded Vegetation (1.39%)	Crops (25.37%)	Scrub (5.47%)	Built Area (9.18%)	Bare Ground (3.49%)	Snow/Ice (2.06%)	OA
DW-Net (Brown et al., 2022)	PA	96.80%	**97.50%**	**60.60%**	68.90%	74.70%	61.70%	95.30%	**92.20%**	**100.00%**	88.40%
	UA	98.60%	87.50%	28.80%	86.00%	97.10%	**64.90%**	**96.70%**	62.90%	78.20%	
	mIoU	/	/	/	/	/	/	/	/	/	
	F-score	0.977	0.922	0.390	0.765	0.844	0.633	0.960	**0.748**	0.878	
SIFDM (this paper)	PA	**98.83%**	96.41%	53.69%	**90.20%**	**90.88%**	**94.33%**	**97.77%**	50.54%	99.89%	**93.39%**
	UA	**99.68%**	**96.71%**	**75.31%**	**86.74%**	**99.41%**	52.82%	96.48%	**97.69%**	**98.27%**	
	mIoU	0.985	0.933	0.457	0.793	0.904	0.512	0.944	0.499	0.982	
	F-score	**0.993**	**0.966**	**0.627**	**0.884**	**0.950**	**0.677**	**0.971**	0.666	**0.991**	

Note: Both models were validated with the same validation samples from Dynamic World training dataset, in which ground truths in validation samples were composited using the voting scheme "Three Expert Strict", that is, where all three expert annotators labeled and all agreed. The accuracy of DW-Net (i.e., the model used to produce the Google Dynamic World LULC product) was collected from the original Dynamic World paper (Brown et al., 2022).

vegetation. However, the ESA WorldCover provides wetland information. We used the flooded vegetation labels to evaluate the accuracy of wetlands for the ESA WorldCover product over GBA for convenience of accuracy evaluation and result comparisons; thus, the accuracy of wetlands is only for the relative comparison. Considering the differences among the four compared products, the coloring scheme for mapping is unified to the same scheme as ESA WorldCover products. The naming of LULC types is also unified as defined in the Dynamic World training dataset. Table 3 shows the accuracy of our annual GBACover products and the other compared LULC products over the GBA from 2019 to 2022. The validation results suggested that GBACover has the highest overall accuracy for the annual LULC mapping in GBA over the four consecutive years of 2019–2022. Meanwhile, Google Dynamic World and Esri Land Cover are the second best for their better performances in separate years. The assessment shows a mean overall accuracy of 87.01% of GBACover for annual mapping in 2019–2022 over GBA and outperforms ESA WorldCover of 83.98% in 2020 and 2021, Esri Land Cover of 85.26%, and Google Dynamic World of 85.06%. If the accuracy assessment for wetlands in ESA WorldCover are excluded due to the absence of validation labels, the mean overall accuracy of ESA WorldCover 2020 and 2021 is 85.82%, which is higher than its actual accuracy because of challenges in the accurate identification of wetlands. It is worth noting that there is temporal variance in the classification accuracy of composited LULC maps across different lengths of periods. Despite the GBACover achieving a mean overall accuracy of 87.01% at an annual scale, the validation of the time series LULC maps generated by the proposed method reports a mean overall accuracy of 80.13% at a near real-time scale. This accuracy was determined by validating LULC maps against manually interpreted samples over the study area.

Table 3
Accuracy comparison of the annual LULC products of the GBA in 2019–2022.

Product Name /Overall Accuracy	2019	2020	2021	2022	Mean
ESA WorldCover (Zanaga et al., 2021)	N/A	83.63%	84.32%	N/A	83.98%
Esri Land Cover (Karra et al., 2021)	84.82%	85.50%	85.98%	84.75%	85.26%
Google Dynamic World (Brown et al., 2022)	85.36%	85.86%	85.10%	83.91%	85.06%
GBACover (this paper)	**87.28%**	**87.65%**	**86.97%**	**86.15%**	**87.01%**

Note: Considering the differences of the LULC types among the products, the manual flooded vegetation labels are used to evaluate the accuracy of wetlands for the ESA WorldCover product; thus, its overall accuracy is only for the relative comparison purpose. The manual validation labels of grass and scrub are merged for accuracy assessment of rangeland defined in Esri Land Cover product. Detailed comparisons of class definitions among the LULC products are provided in Appendix Table 2.

In terms of the LULC mapping accuracy of the different products for specific classes, for a fair comparison with Esri Land Cover, where the grass and scrub are merged into rangeland, we also combine these two classes in other products for the accuracy assessment as shown in Appendix Table 2. The accuracy validation results for the specific classes shown in Fig. 5 indicated that all products acquired overall satisfactory classification results for classes of water, trees, and built area even though GBACover has an overall higher accuracy in the annual LULC maps for most classes with the exception of bare ground among the compared products. Although the classification of flood vegetation, rangeland, and crops is challenging, and the accuracy in terms of F-score for these classes is less satisfactory compared with other classes, the results of GBACover for these hard-classified classes show an obvious improvement, especially for flooded vegetation and rangeland. The annual LULC maps over GBA in the consecutive 4 years from 2019 to 2022 are shown in Fig. 6 for detailed visual comparisons. The comparison over the local area is shown in Fig. 7, which confirms the superiority of GBACover compared with other LUCL products.

4.3. Influence assessment of cloud coverage on LULC mapping

The masks from QA60, Sen2Cor, s2cloudless, and masks generated by the proposed SIFDM model before and after time series refinement are involved in the comparisons to quantitatively evaluate the influence of the accuracy of cloud and cloud shadow masks on LULC mapping. Specifically, the time series LULC maps, generated based on the original Sentinel-2 time series without reconstruction, are validated with the manually labeled samples in GBA. The cloud/cloud shadow contaminated pixels, identified by the compared masks, are excluded from the generated time series LULC maps before validation. This process allows for the evaluation of LULC mapping performances with different masks. The 1188 manually labeled LULC-unchanged sample pixels introduced in Section 2 were used to evaluate the overall accuracy of the pixels' time series LULC classification results, in which each pixel's time series was assumed to remain in the same classes over the study period. The result shows a total of up to 419,861 pixels used for the accuracy evaluations. The number of pixels involved for accuracy evaluation varies with different masks, due to the exclusion of different percentages of invalid cloud/cloud shadow contaminated pixels identified by the compared masks. In addition, two collections of masks generated from the grayscale cloud probability maps of s2cloudless with different binarization thresholds (i.e., 25 and 50) are involved for comparisons. The accuracies of the initial masks generated by SIFDM and their refined masks are separately evaluated for fair comparisons.

The detailed accuracy evaluation results for the time series LULC maps produced using different cloud masks are provided in Table 4, which show significant overall accuracy differences. Specifically, the utilization of the refined and initial cloud masks generated by SIFDM,

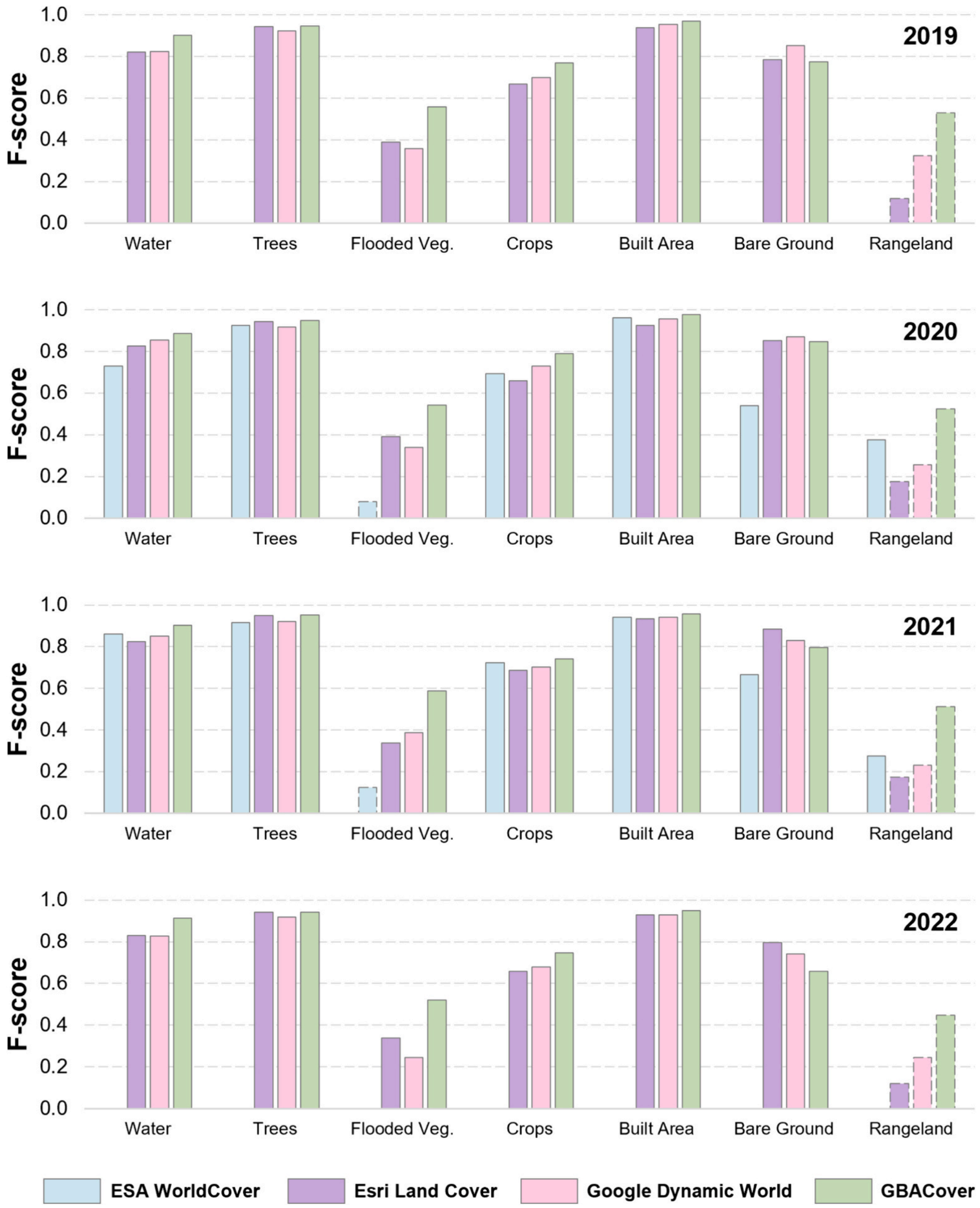

Fig. 5. Accuracy assessment results of the annual LULC products over different classes in 2019–2022. The annual LULC maps of ESA WorldCover in 2019 and 2022 are not available; thus, they are not involved in the comparisons. Meanwhile, the classification accuracy of wetlands in ESA WorldCover 2020 and 2021 is evaluated with the manual flooded vegetation labels and can only be relatively compared. Similar to Esri Land Cover products, the classes of grass and scrub are merged into rangeland for convenience of comparison among different products.

which perform best in cloud and cloud shadow masking compared with other compared masks (Table 1), results in the best overall accuracies of 81.16% and 68.06% for the time series LULC mapping, respectively. The refined masks can even contribute to a higher overall LULC classification accuracy than the original masks generated by SIFDM, confirming the benefits of time series refinement for improving the accuracy of cloud masks. More accurate cloud masks contribute to the higher accuracy of LULC mapping. The two collections of masks generated by s2cloudless result in the overall LULC classification accuracy of 59.35% and 57.04% with two different binarization thresholds of 25 and 50 for the grayscale mask segmentation, respectively. The overall accuracy of the LULC classification with Sen2Cor masks is 64.21%, which is higher than that of s2cloudless, potentially due to the additional cloud shadow information labeled in Sen2Cor masks. The less accurate cloud masks from QA60 result in an overall LULC classification accuracy of 47.99%, which is worth the users' attention for the application of QA60 masks for Sentinel-2 image interpretation.

The above-mentioned results of influence assessments of the cloud coverage provide an intuitive impression of how much the accuracy of the cloud mask affects the accuracy of LULC mapping. However, it is noteworthy that the accuracy of cloud masks is rarely considered in studies relevant to satellite image interpretation based on deep learning (Rußwurm and Körner, 2020; Turkoglu et al., 2021). This oversight raises concerns regarding the accurate mapping of land and water dynamics with optical satellite images, especially in cloudy conditions. This paper not only provides a comprehensive methodology for LULC mapping in cloud-prone areas by incorporating advanced cloud masking and LULC classification models but also assesses the influences of cloud masks on LULC mapping. The assessment reveals that LULC mapping accuracies vary significantly, from 47.99% to 81.16%, when applying

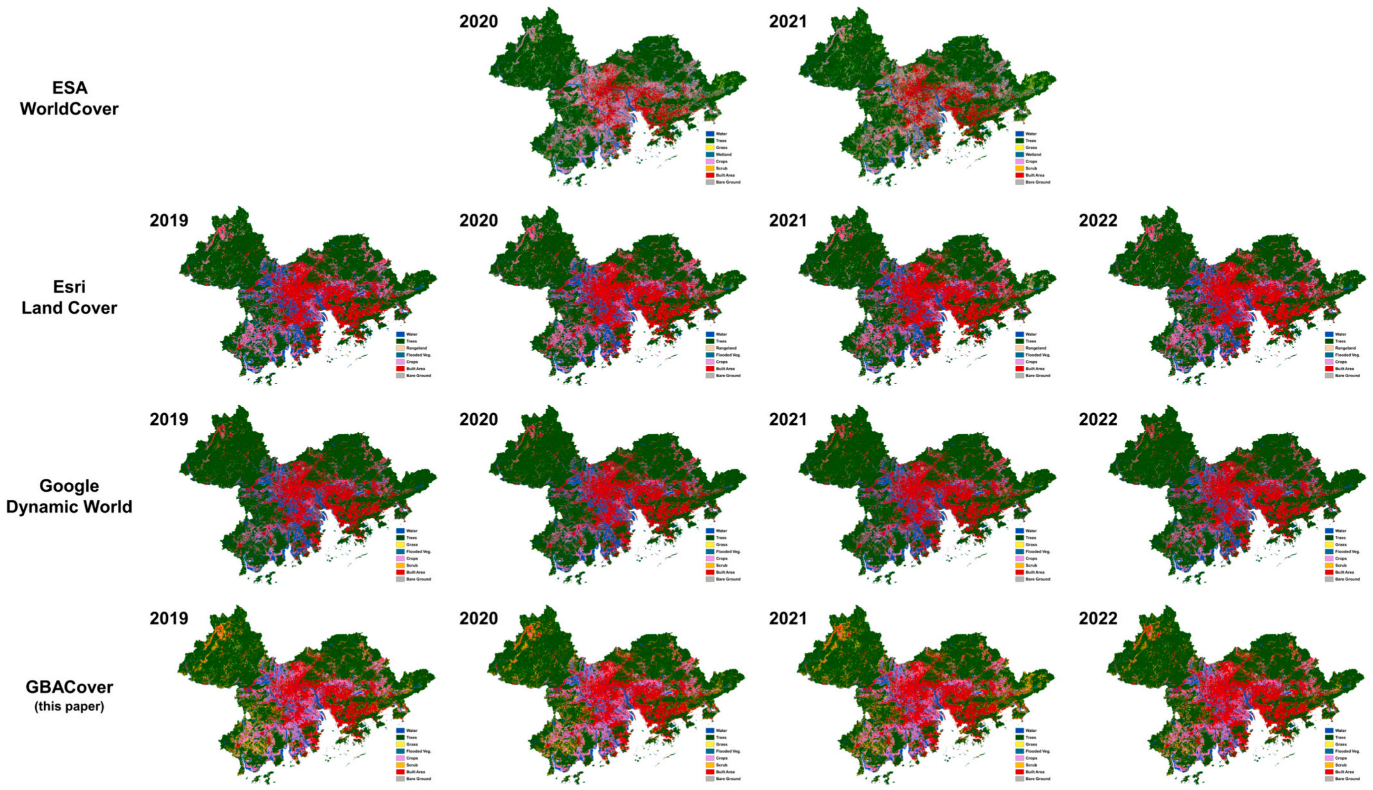

Fig. 6. Comparison of GBACover with other LULC products in 2019–2022.

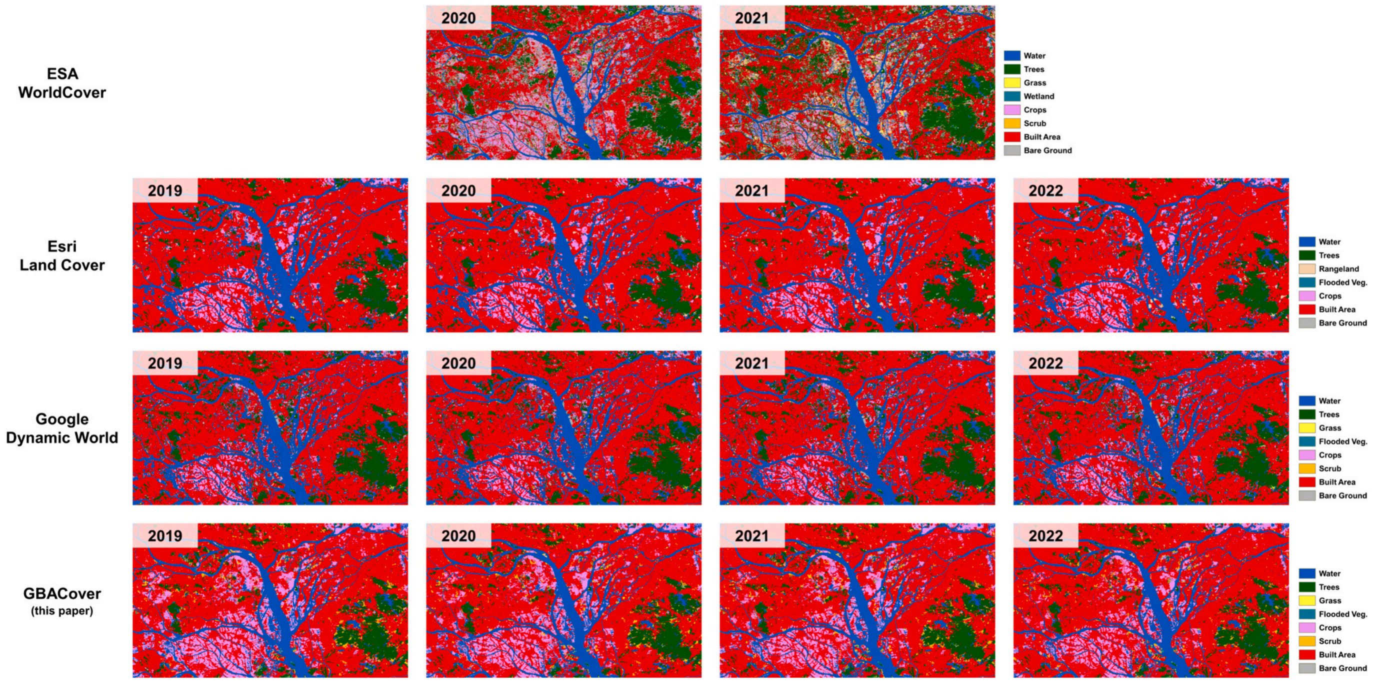

Fig. 7. Comparison of LULC mapping in the Pearl River estuary region in 2019–2022.

different cloud masks, thus underscoring the importance of accurate cloud masks for time series LULC mapping.

5. Discussion

5.1. Benefits of time series reconstruction for LULC mapping

This study applied Whittaker filtering for the time series reconstruction of the contaminated pixels in the image time series, which is expected to benefit dealing with LULC mapping in cloud-prone areas. In

Table 4
Accuracy comparisons of time series LULC mapping using different cloud masks.

Cloud masks used	QA60	Sen2Cor	s2cloudless (>50)	s2cloudless (>25)	SIFDM (Initial)	SIFDM (Refined)
OA of LULC mapping	47.99%	64.21%	57.04%	59.35%	68.06%	81.16%

Note: Two collections of masks generated from grayscale cloud probability maps of s2cloudless with different binarization thresholds are involved for comparisons. The accuracies of initial masks generated by SIFDM and their refined masks are evaluated separately for fair comparisons.

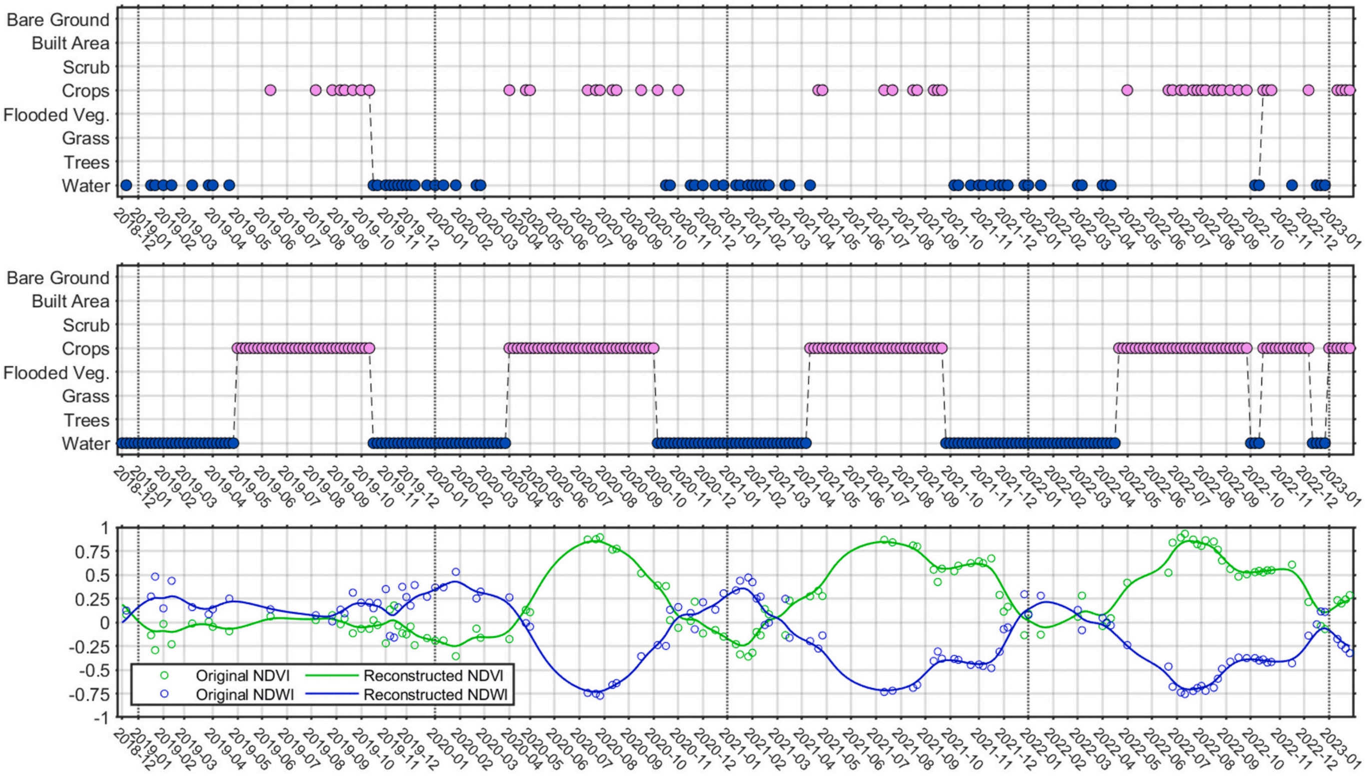

Fig. 8. Comparisons of time series mapping with the original and the reconstructed cloud-free images. The upper figure denotes the results generated with the original images, in which the missing points in the time series are caused by cloud coverage. The middle figure refers to the time series mapping results with the reconstructed images and after post-classification processing. The lower figure provides the NDVI and NDWI time series derived from both the original and the reconstructed cloud-free images, respectively.

Fig. 8, taking LULC mapping in agriculture land as an example, an agriculture land can either be covered by water or planted with crops, and the land cover change between water and crops may occur several times within a year. In this case, the mapping results vary with the involved images for LULC mapping, resulting in biases for mapping of LULC types that involve land and water interactions (e.g., crops and wetland). The accurate mapping of the LULC types that involve land and water interactions becomes more challenging because the cloud coverage on the image time series randomly occurs. Reconstruction of the cloud- or cloud shadow-contaminated areas in the image time series is necessary for the accurate LULC mapping to minimize the biases and errors in time series mapping caused by the random cloud coverage and selected images. Composting high-quality annual, seasonal, and even monthly LUCL maps is possible because of the reconstructed cloud-free image time series. The time series change patterns can be considered to refine the mapping results.

5.2. Annual LULC mapping with dense image time series

Existing annual LULC products are typically produced based on images acquired during the vegetation growing season (Chen et al., 2015; Yang and Huang, 2021). However, optical satellite imagery suffers from dense cloud coverage during these rainy and cloudy seasons, resulting in limited available cloud-free images for the annual LULC mapping. Accordingly, the annual mapping results are generated only based on one or several valid satellite observations. In this study, the annual LULC maps are generated and composited based on all available Sentinel-2 images for a year. Post-classification processing for the time series LULC maps is additionally conducted to improve the time series consistency and filter out noises. Consequently, the accuracy and robustness of the annual LULC mapping can be improved and enhanced. LULC mapping with dense image time series holds promise in capturing the periodic LULC change patterns (e.g., the land and water interactions in crop areas), which can be identified through the time series post-classification processing and analysis. In Fig. 9, the frequencies of the LULC types can be acquired from dense time series LULC maps, from which the annual LULC change trends can be quantitatively measured with high accuracy and in an intuitive manner (e.g., the change trends from crops to built area over the 4 years). Therefore, dense time series post-classification processing and analysis benefit LULC mapping in terms of accuracy and robustness.

5.3. Limitations

Although the proposed time series LULC mapping method achieved better performances than the compared methods in terms of cloud masking and LULC classification, the limitations still exist with the proposed method. On the one hand, the performance of deep models is

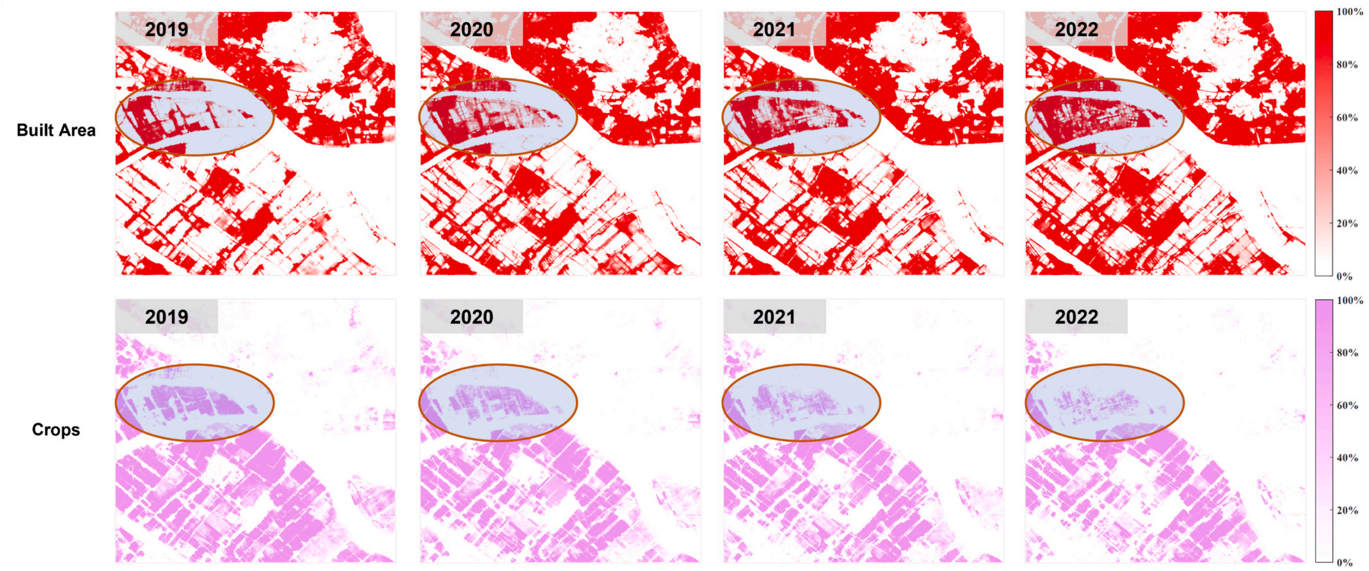

Fig. 9. Example of identifying change trends from the annual LULC frequency maps. The transition from crops to built area in this example can be clearly interpreted from their corresponding annual LULC frequency maps, which are generated based on dense time series LULC maps and represent the proportion of times a category is detected out of the total number of observations in a year.

subjective to the amount and quality of training samples used in this study, especially the Dynamic World training dataset for the training of the LULC classification model, which is rough and lacks details in object boundaries, thus limiting the model performance in identifying slim ground objects with minor sizes, such as bridges and roads, and leaving space for further improvements with high-quality training labels. Meanwhile, the performance of deep models can be further enhanced by using state-of-the-art deep architectures, such as vision transformer based foundation models (Cha et al., 2023; Wang and Mountrakis, 2023), which have been proven effective in image interpretation, to exploit the potentials of the proposed method. On the other hand, cumulated processing errors can occur because multiple steps are involved in the LULC method, especially the accuracy of masks, time series reconstruction, and LULC classification, as exemplified in Fig. 10. Additionally, short-term LULC changes might be hidden by clouds and cannot be effectively reconstructed and captured or might be smoothened in post-classification processing, which will result in a decrease in accuracy for near real-time and monthly LULC mapping.

6. Conclusions

An integrated method for LULC mapping using dense Sentinel-2 time series is proposed in this paper. This method can be used for near real-time, monthly, seasonal, and annual mapping in cloud-prone areas, despite the temporal variance in classification accuracy across different lengths of periods. The proposed methods have shown their superiority over the compared methods in cloud masking and LULC classification by developing deep models for improving the accuracy of cloud masking and LULC classification, employing a time series reconstruction method for filling cloud-contaminated pixels, and applying time series post-classification processing and analysis. The application of the proposed method in the GBA suggested that it can generate accurate LULC maps, achieving a mean overall of 87.01% at an annual scale and 80.13% at a near real-time scale, thereby outperforming the compared annual LULC products. We evaluated the influence of cloud coverage on LULC mapping, suggesting the necessity of developing advanced cloud masking methods to improve LULC mapping accuracy, as has been done in this study. The benefits of time series reconstruction and LULC mapping with dense time series images were also discussed, which illustrate their contributions to LULC mapping, especially in improving mapping results for LULC types involved in land and water interactions and in cloud-prone areas.

The deep models in the proposed method are trained on datasets collected globally, which can be used for LULC mapping in other regions worldwide beyond the study area. Meanwhile, the proposed method can also be applied for near real-time monitoring of a single LULC category (e.g., time series monitoring of cropland, wetland, and inundated land), thus having a broad range of potential applications. Nevertheless, the limitations of the proposed method in the performance of deep models and error propagation brought by multiple processing steps leave much space for further improvements, such as integrating with the state-of-the-art large foundation model and introducing quality control during the production of LULC maps. Additionally, field surveys in the study area are required to further validate the LULC products, especially in challenging mapping scenarios, such as for grass/shrub, wetland, and flood vegetation classes. In the future, with the introduction of more advanced deep models (e.g., large foundation models) and multi-modal data (e.g., combination of Sentinel 1 and 2), the LULC mapping with dense image time series in cloud- and rain-prone areas can be further enhanced. The potentials of SAR images can be fully exploited to benefit the identification of dynamic land change patterns during periods of persistent cloud coverages.

CRediT authorship contribution statement

Zhiwei Li: Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Qihao Weng:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Yuhan Zhou:** Writing – original draft, Visualization, Methodology, Data curation, Conceptualization. **Peng Dou:** Data curation, Conceptualization. **Xiaoli Ding:** Resources, Project administration.

Declaration of competing interest

None.

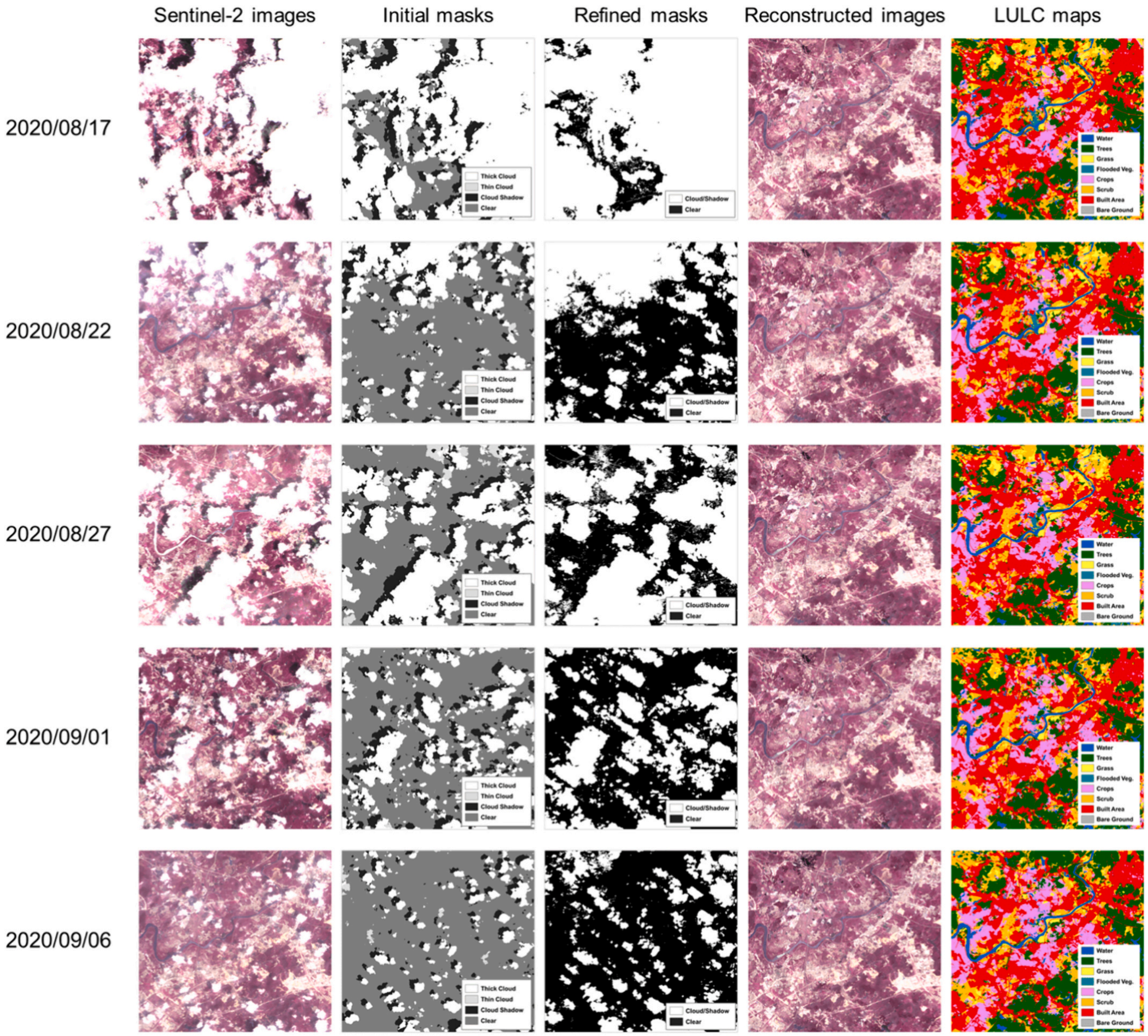

Fig. 10. Example of intermediate results of the proposed LULC mapping method.

Data availability

Data will be made available on request.

Acknowledgments

We acknowledge the research support of the Global STEM Professorship by the Hong Kong SAR Government and The Hong Kong Polytechnic University Start-up Fund, and the financial support by the Institute of Land and Space at The Hong Kong Polytechnic University. This research was partially supported by the National Natural Science Foundation of China (No. 42101357). We also thank Yujun Mao at Wuhan University and Longjie Ye at The Hong Kong Polytechnic University for their assistance in collecting the experimental data. Additionally, we extend our gratitude to the editors and four anonymous reviewers, whose constructive comments have helped to improve the quality of this paper.

Appendix A

Appendix Table 1
The network structure of the SIFDM model.

Module	Unit level	Layers	Kernel size	Output size	Remarks
Dual encoders	Level 1	2*(Conv + BN + ReLU)	3 × 3	(512 × 512) × 64	
	Level 2	Max Pooling	2 × 2	(256 × 256) × 64	
		2*(Conv + BN + ReLU)	3 × 3	(256 × 256) × 128	
	Level 3	Max Pooling	2 × 2	(128 × 128) × 128	
		2*(Conv + BN + ReLU)	3 × 3	(128 × 128) × 256	
	Level 4	Max Pooling	2 × 2	(64 × 64) × 256	
		2*(Conv + BN + ReLU)	3 × 3	(64 × 64) × 512	
	Level 5	Max Pooling	2 × 2	(32 × 32) × 512	
		2*(Conv + BN + ReLU)	3 × 3	(32 × 32) × 512	
Decoder	Level 6	Concat + (Conv + BN + ReLU)	1 × 1	(32 × 32) × 512	Concat Level 5 features from dual encoders
	Level 7	Concat + (Conv + BN + ReLU)	1 × 1	(64 × 64) × 512	Concat Level 4 features from dual encoders
	Level 8	Concat + (Conv + BN + ReLU)	1 × 1	(128 × 128) × 256	Concat Level 3 features from dual encoders
	Level 9	Concat + (Conv + BN + ReLU)	1 × 1	(256 × 256) × 128	Concat Level 2 features from dual encoders
	Level 10	Concat + (Conv + BN + ReLU)	1 × 1	(512 × 512) × 64	Concat Level 1 features from dual encoders
	Level 11	Bilinear Upsampling + Concat	–	(64 × 64) × 1024	Upsample Level 6 features, then Concat them with Level 7 features
		2*(Conv + BN + ReLU)	3 × 3	(64 × 64) × 256	
	Level 12	Bilinear Upsampling + Concat	–	(128 × 128) × 512	Upsample Level 11 features, then Concat them with Level 8 features
		2*(Conv + BN + ReLU)	3 × 3	(128 × 128) × 128	
	Level 13	Bilinear Upsampling + Concat	–	(256 × 256) × 256	Upsample Level 12 features, then Concat them with Level 9 features
		2*(Conv + BN + ReLU)	3 × 3	(256 × 256) × 64	
	Level 14	Bilinear Upsampling + Concat	–	(512 × 512) × 128	Upsample Level 13 features, then Concat them with Level 10 features
		2*(Conv + BN + ReLU)	3 × 3	(512 × 512) × 64	
	Level 15	Conv	1 × 1	(512 × 512) × N	N is the number of classes in output

Note: The network structure is provided using input images with a height and width of 512 × 512 as an example. The stride size is set to 1 in all relevant operations. The padding size in all relevant operations is adaptively set to keep the size of the feature maps unchanged. Conv: Convolution. BN: Batch Normalization.

Appendix Table 2
Comparisons of class definitions among LULC products (Adapted from Wang and Mountrakis, 2023).

ESA WorldCover	Esri Land Cover	Google Dynamic World	GBACover
Permanent water bodies *This class includes any geographic area covered for most of the year (>9 months) by water bodies, e.g., lakes, reservoirs and rivers. Can be either fresh or salt-water bodies. In some cases, the water can be frozen for part of the year (<9 months).*	Water *Areas where water was predominantly present throughout the year. May not cover areas with sporadic or ephemeral water and contains little to no sparse vegetation, no rock outcrop nor built up features like docks.*	Water *Water is present in the image. Contains little-to-no sparse vegetation, no rock outcrop, and no built-up features like docks. Does not include land that can or has previously been covered by water.*	
Tree cover *This class includes any geographic area dominated by trees with a cover of 10% or more. Other land cover classes (shrubs and/or herbs in the understory, built-up, permanent water bodies, etc.) can be present below the canopy, even with a density higher than trees. Areas planted with trees for afforestation purposes and plantations (e.g., oil palm, olive trees) are included in this class. This class also includes tree-covered areas seasonally or permanently flooded with fresh water except for mangroves.*	Trees *Any significant clustering of tall (~15 ft or higher) dense vegetation, typically with a closed or dense canopy (i.e., dense/tall vegetation with ephemeral water or canopy too thick to detect water underneath).*	Trees *Any significant clustering of dense vegetation, typically with a closed or dense canopy. Taller and darker than surrounding vegetation (if surrounded by other vegetation).*	
Grassland *This class includes any geographic area (e.g., grasslands, prairies, steppes, savannahs, pastures) dominated by natural herbaceous plants (i.e., plants without persistent stem or shoots above ground and lacking definite firm structure) with a cover of 10% or more, irrespective of different human and/or animal activities, such as grazing, selective fire management etc. Woody plants (trees and/or shrubs) can be present assuming their cover is <10%. It may also contain uncultivated cropland areas (without harvest/ bare soil period) in the reference year.*	Rangeland *Open areas covered in homogenous grasses with little to no taller vegetation; wild cereals and grasses with no obvious human plotting (i.e., not a plotted field). Mix of small clusters of plants or single plants dispersed on a landscape that shows exposed soil or rock; scrub-filled clearings within dense forests that are clearly not taller than trees.*	Grass *Open areas covered in homogenous grasses with little to no taller vegetation. Other homogenous areas of grass-like vegetation (blade-type leaves) that appear different from trees and shrubland. Wild cereals and grasses with no obvious human plotting (i.e. not a structured field).*	
Shrubland *This class includes any geographic area dominated by natural shrubs having a cover of 10% or more. Shrubs are defined as woody perennial plants with persistent and woody stems and without any defined main stem being <5 m tall. Trees can be present in scattered form if their cover is <10%. Herbaceous plants can also be present at any density. The shrub foliage can be either evergreen or deciduous.*		Shrub & Scrub *Mix of small clusters of plants or individual plants dispersed on a landscape that shows exposed soil and rock. Scrub-filled clearings within dense forests that are clearly not taller than trees. Appear grayer/browner due to less dense leaf cover.*	

(continued on next page)

Appendix Table 2 (*continued*)

ESA WorldCover	Esri Land Cover	Google Dynamic World	GBACover
Herbaceous Wetland *Land dominated by natural herbaceous vegetation (cover of 10% or more) that is permanently or regularly flooded by fresh, brackish or salt water. It excludes unvegetated sediment and swamp forests (classified as tree cover) and mangroves).*	Flooded vegetation *Areas of any type of vegetation with obvious intermixing of water throughout a majority of the year. Seasonally flooded area that is a mix of grass/shrub/trees/bare ground.*	Flooded vegetation *Areas of any type of vegetation with obvious intermixing of water. Do not assume an area is flooded if flooding is observed in another image. Seasonally flooded areas that are a mix of grass/shrub/trees/bare ground.*	
Cropland *Land covered with annual cropland that is sowed/planted and harvestable at least once within the 12 months after the sowing/planting date. The annual cropland produces an herbaceous cover and is sometimes combined with some tree or woody vegetation. Note that perennial woody crops will be classified as the appropriate tree cover or shrub land cover type. Greenhouses are considered as built-up.*	Crops *Human planted/plotted cereals, grasses, and crops not at tree height.*	Crops *Human planted/plotted cereals, grasses, and crops.*	
Built-up *Land covered by buildings, roads and other man-made structures such as railroads. Buildings include both residential and industrial buildings. Urban green (e.g., parks, sport facilities) is not included in this class. Waste dump deposits and extraction sites are considered as bare.*	Built area *Human made structures and major road and rail networks. Large homogenous impervious surfaces including parking structures, office buildings, and residential housing.*	Built area *Clusters of human-made structures or individual very large human-made structures. Contained industrial, commercial, and private building, and the associated parking lots. A mixture of residential buildings, streets, lawns, trees, isolated residential structures or buildings surrounded by vegetative land covers. Major road and rail networks outside of the predominant residential areas. Large homogeneous impervious surfaces, including parking structures, large office buildings, and residential housing developments containing clusters of cul-de-sacs.*	
Bare/sparse vegetation *Lands with exposed soil, sand or rocks, and never has >10% vegetated cover during any time of the year.*	Bare ground *Areas of rock or soil with very sparse to no vegetation for the entire year. Large areas of sand and deserts with no to little vegetation.*	Bare ground *Areas of rock or soil containing very sparse to no vegetation. Large areas of sand and deserts with no to little vegetation. Large individual or dense networks of dirt roads.*	

References

Aybar, C., Ysuhuaylas, L., Loja, J., Gonzales, K., Herrera, F., Bautista, L., Yali, R., Flores, A., Diaz, L., Cuenca, N., Espinoza, W., Prudencio, F., Llactayo, V., Montero, D., Sudmanns, M., Tiede, D., Mateo-García, G., Gómez-Chova, L., 2022. CloudSEN12, a global dataset for semantic understanding of cloud and cloud shadow in Sentinel-2. Sci. Data 9. https://doi.org/10.1038/s41597-022-01878-2.

Baetens, L., Desjardins, C., Hagolle, O., 2019. Validation of copernicus Sentinel-2 cloud masks obtained from MAJA, Sen2Cor, and FMask processors using reference cloud masks generated with a supervised active learning procedure. Remote Sens. 11, 1–25. https://doi.org/10.3390/rs11040433.

Blaschke, T., 2010. Object based image analysis for remote sensing. ISPRS J. Photogramm. Remote Sens. 65, 2–16. https://doi.org/10.1016/j.isprsjprs.2009.06.004.

Brown, C.F., Brumby, S.P., Guzder-Williams, B., Birch, T., Hyde, S.B., Mazzariello, J., Czerwinski, W., Pasquarella, V.J., Haertel, R., Ilyushchenko, S., Schwehr, K., Weisse, M., Stolle, F., Hanson, C., Guinan, O., Moore, R., Tait, A.M., 2022. Dynamic world, near real-time global 10 m land use land cover mapping. Sci. Data 9, 251. https://doi.org/10.1038/s41597-022-01307-4.

Cha, K., Seo, J., Lee, T., 2023. A Billion-scale Foundation Model for Remote Sensing Images (arXiv Prepr. arXiv2304.05215).

Chambrelan, A., 2012. Sentinel-2 MSI - Level-1 Algorithm Theoretical Bases Document. European Space Agency.

Chen, Jun, Chen, Jin, Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., Lu, M., Zhang, W., Tong, X., Mills, J., 2015. Global land cover mapping at 30 m resolution: a POK-based operational approach. ISPRS J. Photogramm. Remote Sens. 103, 7–27. https://doi.org/10.1016/j.isprsjprs.2014.09.002.

Cheng, B., Schwing, A.G., Kirillov, A., 2021. Per-pixel classification is not all you need for semantic segmentation. Adv. Neural Inf. Proces. Syst. 22, 17864–17875.

Coluzzi, R., Imbrenda, V., Lanfredi, M., Simoniello, T., 2018. A first assessment of the Sentinel-2 level 1-C cloud mask product to support informed surface analyses. Remote Sens. Environ. 217, 426–443. https://doi.org/10.1016/j.rse.2018.08.009.

Domnich, M., Sünter, I., Trofimov, H., Wold, O., Harun, F., Kostiukhin, A., Järveoja, M., Veske, M., Tamm, T., Voormansik, K., Olesk, A., Boccia, V., Longepe, N., Cadau, E.G., 2021. KappaMask: AI-based cloudmask processor for Sentinel-2. Remote Sens. 13, 4140.

Dou, P., Shen, H., Li, Z., Guan, X., 2021. Time series remote sensing image classification framework using combination of deep learning and multiple classifiers system. Int. J. Appl. Earth Obs. Geoinf. 103, 102477 https://doi.org/10.1016/j.jag.2021.102477.

Duro, D.C., Franklin, S.E., Dubé, M.G., 2012. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. Remote Sens. Environ. 118, 259–272. https://doi.org/10.1016/j.rse.2011.11.020.

Eilers, P.H.C., 2003. A perfect smoother. Anal. Chem. 75, 3631–3636.

Fan, R., Feng, R., Wang, L., Yan, J., Zhang, X., 2020. Semi-MCNN: a semisupervised multi-CNN ensemble learning method for urban land cover classification using submeter HRRS images. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 4973–4987. https://doi.org/10.1109/JSTARS.2020.3019410.

Foody, G.M., 1995. Land cover classification by an artificial neural network with ancillary information. Int. J. Geogr. Inf. Syst. 9, 527–542. https://doi.org/10.1080/02693799508902054.

Frantz, D., Haß, E., Uhl, A., Stoffels, J., Hill, J., 2018. Improvement of the Fmask algorithm for Sentinel-2 images: separating clouds from bright surfaces based on parallax effects. Remote Sens. Environ. 215, 471–481. https://doi.org/10.1016/j.rse.2018.04.046.

Fu, G., Liu, C., Zhou, R., Sun, T., Zhang, Q., 2017. Classification for high resolution remote sensing imagery using a fully convolutional network. Remote Sens. 9, 498. https://doi.org/10.3390/rs9050498.

Gaetano, R., Ienco, D., Ose, K., Cresson, R., 2018. A two-branch CNN architecture for land cover classification of PAN and MS imagery. Remote Sens. 10, 1746. https://doi.org/10.3390/rs10111746.

Gómez, C., White, J.C., Wulder, M.A., 2016. Optical remotely sensed time series data for land cover classification: a review. ISPRS J. Photogramm. Remote Sens. 116, 55–72. https://doi.org/10.1016/j.isprsjprs.2016.03.008.

Gong, P., Liu, H., Zhang, M., Li, C., Wang, J., Huang, H., Clinton, N., Ji, L., Li, Wenyu, Bai, Y., Chen, B., Xu, B., Zhu, Z., Yuan, C., Ping Suen, H., Guo, J., Xu, N., Li, Weijia, Zhao, Y., Yang, J., Yu, C., Wang, X., Fu, H., Yu, L., Dronova, I., Hui, F., Cheng, X., Shi, X., Xiao, F., Liu, Q., Song, L., 2019. Stable classification with limited sample: transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017. Sci. Bull. 64, 370–373. https://doi.org/10.1016/j.scib.2019.03.002.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earth engine: planetary-scale geospatial analysis for everyone. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2017.06.031.

Hagolle, O., Huc, M., Desjardins, C., Auer, S., Richter, R., 2017. MAJA Algorithm Theoretical Basis Document. https://doi.org/10.5281/zenodo.1209633.

Hong, D., Gao, L., Yokoya, N., Yao, J., Chanussot, J., Du, Q., Zhang, B., 2020. More diverse means better: multimodal deep learning meets remote-sensing imagery classification. IEEE Trans. Geosci. Remote Sens. 59, 4340–4354.

Huang, B., Zhao, B., Song, Y., 2018. Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery. Remote Sens. Environ. 214, 73–86. https://doi.org/10.1016/j.rse.2018.04.050.

Ienco, D., Interdonato, R., Gaetano, R., Ho Tong Minh, D., 2019. Combining Sentinel-1 and Sentinel-2 satellite image time series for land cover mapping via a multi-source deep learning architecture. ISPRS J. Photogramm. Remote Sens. 158, 11–22. https://doi.org/10.1016/j.isprsjprs.2019.09.016.

Interdonato, R., Ienco, D., Gaetano, R., Ose, K., 2019. DuPLO: a DUal view point deep learning architecture for time series classificatiOn. ISPRS J. Photogramm. Remote Sens. 149, 91–104. https://doi.org/10.1016/j.isprsjprs.2019.01.011.

Karra, K., Kontgis, C., Statman-Weil, Z., Mazzariello, J.C., Mathis, M., Brumby, S.P., 2021. Global land use / land cover with Sentinel 2 and deep learning. In: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. IEEE, pp. 4704–4707. https://doi.org/10.1109/IGARSS47720.2021.9553499.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., Dollár, P., Girshick, R., 2023. Segment Anything (arXiv Prepr. arXiv2304.02643).

Kussul, N., Lavreniuk, M., Skakun, S., Shelestov, A., 2017. Deep learning classification of land cover and crop types using remote sensing data. IEEE Geosci. Remote Sens. Lett. 14, 778–782. https://doi.org/10.1109/LGRS.2017.2681128.

Li, W., Fu, H., Yu, L., Gong, P., Feng, D., Li, C., Clinton, N., 2016. Stacked autoencoder-based deep learning for remote-sensing image classification: a case study of African land-cover mapping. Int. J. Remote Sens. 37, 5632–5646. https://doi.org/10.1080/01431161.2016.1246775.

Li, Z., Shen, H., Li, H., Xia, G., Gamba, P., Zhang, L., 2017. Multi-feature combined cloud and cloud shadow detection in GaoFen-1 wide field of view imagery. Remote Sens. Environ. 191, 342–358. https://doi.org/10.1016/j.rse.2017.01.026.

Li, Y., Zhang, H., Xue, X., Jiang, Y., Shen, Q., 2018. Deep learning for remote sensing image classification: a survey. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 8, e1264 https://doi.org/10.1002/widm.1264.

Li, J., Wu, Z., Hu, Z., Jian, C., Luo, S., Mou, L., Zhu, X.X., Molinier, M., 2021. A lightweight deep learning-based cloud detection method for sentinel-2A imagery fusing multiscale spectral and spatial features. IEEE Trans. Geosci. Remote Sens. 60, 5401219. https://doi.org/10.1109/TGRS.2021.3069641.

Li, Z., Shen, H., Weng, Q., Zhang, Y., Dou, P., Zhang, L., 2022a. Cloud and cloud shadow detection for optical satellite imagery: features, algorithms, validation, and prospects. ISPRS J. Photogramm. Remote Sens. 188, 89–108. https://doi.org/10.1016/j.isprsjprs.2022.03.020.

Li, Y., Zhou, Y., Zhang, Y., Zhong, L., Wang, J., Chen, J., 2022b. DKDFN: domain knowledge-guided deep collaborative fusion network for multimodal unitemporal remote sensing land cover classification. ISPRS J. Photogramm. Remote Sens. 186, 170–189. https://doi.org/10.1016/j.isprsjprs.2022.02.013.

Li, Z., He, W., Cheng, M., Hu, J., Yang, G., Zhang, H., 2023. SinoLC-1: the first 1-meter resolution national-scale land-cover map of China created with the deep learning framework and open-access data. Earth Syst. Sci. Data Discuss. 7707461.

Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollar, P., 2020. Focal loss for dense object detection. IEEE Trans. Pattern Anal. Mach. Intell. 318–327. https://doi.org/10.1109/TPAMI.2018.2858826.

Ling, J., Zhang, H., Lin, Y., 2021. Improving urban land cover classification in cloud-prone areas with polarimetric Sar images. Remote Sens. 13, 4708. https://doi.org/10.3390/rs13224708.

Liu, S., Qi, Z., Li, X., Yeh, A.G.O., 2019. Integration of convolutional neural networks and object-based post-classification refinement for land use and land cover mapping with optical and Sar data. Remote Sens. 11, 690. https://doi.org/10.3390/rs11060690.

López-Puigdollers, D., Mateo-García, G., Gómez-Chova, L., 2021. Benchmarking deep learning models for cloud detection in landsat-8 and sentinel-2 images. Remote Sens. 13, 1–20. https://doi.org/10.3390/rs13050992.

Ma, L., Li, M., Ma, X., Cheng, L., Du, P., Liu, Y., 2017. A review of supervised object-based land-cover image classification. ISPRS J. Photogramm. Remote Sens. 130, 277–293. https://doi.org/10.1016/j.isprsjprs.2017.06.001.

Milletari, F., Navab, N., Ahmadi, S.A., 2016. V-net: fully convolutional neural networks for volumetric medical image segmentation. In: Proceedings - 2016 4th International Conference on 3D Vision, 3DV 2016. IEEE, pp. 565–571. https://doi.org/10.1109/3DV.2016.79.

Myint, S.W., Gober, P., Brazel, A., Grossman-Clarke, S., Weng, Q., 2011. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. Remote Sens. Environ. 115, 1145–1161. https://doi.org/10.1016/j.rse.2010.12.017.

OSM, 2017. OpenStreetMap. Planet dump. retrieved from. https://planet.osm.org.

Qi, Z., Yeh, A.G.O., Li, X., Lin, Z., 2012. A novel algorithm for land use and land cover classification using RADARSAT-2 polarimetric SAR data. Remote Sens. Environ. 118, 21–39. https://doi.org/10.1016/j.rse.2011.11.001.

Qiu, C., Mou, L., Schmitt, M., Zhu, X.X., 2019a. Local climate zone-based urban land cover classification from multi-seasonal Sentinel-2 images with a recurrent residual network. ISPRS J. Photogramm. Remote Sens. 154, 151–162. https://doi.org/10.1016/j.isprsjprs.2019.05.004.

Qiu, S., Zhu, Z., He, B., 2019b. Fmask 4.0: improved cloud and cloud shadow detection in Landsats 4–8 and Sentinel-2 imagery. Remote Sens. Environ. 231, 111205 https://doi.org/10.1016/j.rse.2019.05.024.

Richter, R., Louis, J., Berthelot, B., 2012. Sentinel-2 MSI – Level 2A Products Algorithm Theoretical Basis Document, 49. European Space Agency, (Special Publication) ESA SP, pp. 1–72.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.

Rußwurm, M., Körner, M., 2020. Self-attention for raw optical satellite time series classification. ISPRS J. Photogramm. Remote Sens. 169, 421–435. https://doi.org/10.1016/j.isprsjprs.2020.06.006.

Sanchez, A.H., Picoli, M.C.A., Camara, G., Andrade, P.R., Chaves, M.E.D., Lechler, S., Soares, A.R., Marujo, R.F.B., Simões, R.E.O., Ferreira, K.R., Queiroz, G.R., 2020. Comparison of cloud cover detection algorithms on sentinel-2 images of the Amazon tropical forest. Remote Sens. 12, 1–16. https://doi.org/10.3390/RS12081284.

Scott, G.J., Marcum, R.A., Davis, C.H., Nivin, T.W., 2017. Fusion of deep convolutional neural networks for land cover classification of high-resolution imagery. IEEE Geosci. Remote Sens. Lett. 14, 1638–1642. https://doi.org/10.1109/LGRS.2017.2722988.

Song, H., Yang, Y., Gao, X., Zhang, M., Li, S., Liu, B., Wang, Y., Kou, Y., 2023. Joint classification of hyperspectral and LiDAR data using binary-tree transformer network. Remote Sens. 15, 1–16. https://doi.org/10.3390/rs15112706.

Tait, A.M., Brumby, S.P., Hyde, S.B., Mazzariello, J., Corcoran, M., 2021. Dynamic world training dataset for global land use and land cover categorization of satellite imagery. PANGAEA. https://doi.org/10.1594/PANGAEA 933475.

Talukdar, S., Singha, P., Mahato, S., Pal, S., Liou, Y.-A., Rahman, A., 2020. Land-use land-cover classification by machine learning classifiers for satellite observations—a review. Remote Sens. 12, 1135.

Tarrio, K., Tang, X., Masek, J.G., Claverie, M., Ju, J., Qiu, S., Zhu, Z., Woodcock, C.E., 2020. Comparison of cloud detection algorithms for Sentinel-2 imagery. Sci. Remote Sens. 2, 100010 https://doi.org/10.1016/j.srs.2020.100010.

Tong, X.-Y., Xia, G.-S., Lu, Q., Shen, H., Li, S., You, S., Zhang, L., 2020. Land-cover classification with high-resolution remote sensing images using transferable deep models. Remote Sens. Environ. 237, 111322.

Tong, X.Y., Xia, G.S., Zhu, X.X., 2023. Enabling country-scale land cover mapping with meter-resolution satellite imagery. ISPRS J. Photogramm. Remote Sens. 196, 178–196. https://doi.org/10.1016/j.isprsjprs.2022.12.011.

Tracewski, L., Bastin, L., Fonte, C.C., 2017. Repurposing a deep learning network to filter and classify volunteered photographs for land cover and land use characterization. Geo-spat. Inf. Sci. 20, 252–268. https://doi.org/10.1080/10095020.2017.1373955.

Turkoglu, M.O., D'Aronco, S., Perich, G., Liebisch, F., Streit, C., Schindler, K., Wegner, J. D., 2021. Crop mapping from image time series: deep learning with multi-scale label hierarchies. Remote Sens. Environ. 264, 112603 https://doi.org/10.1016/j.rse.2021.112603.

Vali, A., Comai, S., Matteucci, M., 2020. Deep learning for land use and land cover classification based on hyperspectral and multispectral earth observation data: a review. Remote Sens. 12, 2495.

Waleed, M., Mubeen, M., Ahmad, A., Habib-ur-Rahman, M., Amin, A., Farid, H.U., Hussain, S., Ali, M., Qaisrani, S.A., Nasim, W., Javeed, H.M.R., Masood, N., Aziz, T., Mansour, F., EL Sabagh, A., 2022. Evaluating the efficiency of coarser to finer resolution multispectral satellites in mapping paddy rice fields using GEE implementation. Sci. Rep. 12, 13210. https://doi.org/10.1038/s41598-022-17454-y.

Wang, Z., Mountrakis, G., 2023. Accuracy assessment of eleven medium resolution global and regional land cover land use products: a case study over the conterminous United States. Remote Sens. 15, 3186. https://doi.org/10.3390/rs15123186.

Wang, L., Li, R., Zhang, C., Fang, S., Duan, C., Meng, X., Atkinson, P.M., 2022. UNetFormer: a UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery. ISPRS J. Photogramm. Remote Sens. 190, 196–214. https://doi.org/10.1016/j.isprsjprs.2022.06.008.

Weng, Q., 2002. Land use change analysis in the Zhujiang Delta of China using satellite remote sensing, GIS and stochastic modelling. J. Environ. Manag. 64, 273–284. https://doi.org/10.1006/jema.2001.0509.

Whittaker, E.T., 1922. On a new method of graduation. Proc. Edinb. Math. Soc. 41, 63–75. https://doi.org/10.1017/s0013091500077853.

Xing, H., Meng, Y., Wang, Z., Fan, K., Hou, D., 2018. Exploring geo-tagged photos for land cover validation with deep learning. ISPRS J. Photogramm. Remote Sens. 141, 237–251. https://doi.org/10.1016/j.isprsjprs.2018.04.025.

Xu, G., Zhu, X., Fu, D., Dong, J., Xiao, X., 2017. Automatic land cover classification of geo-tagged field photos by deep learning. Environ. Model Softw. 91, 127–134. https://doi.org/10.1016/j.envsoft.2017.02.004.

Yan, G., Mas, J.F., Maathuis, B.H.P., Xiangmin, Z., Van Dijk, P.M., 2006. Comparison of pixel-based and object-oriented image classification approaches - a case study in a coal fire area, Wuda, Inner Mongolia, China. Int. J. Remote Sens. 27, 4039–4055. https://doi.org/10.1080/01431160600702632.

Yang, J., Huang, X., 2021. The 30m annual land cover dataset and its dynamics in China from 1990 to 2019. Earth Syst. Sci. Data 13, 3907–3925. https://doi.org/10.5194/essd-13-3907-2021.

Yang, Y., Huang, Q., Wu, W., Luo, J., Gao, L., Dong, W., Wu, T., Hu, X., 2017. Geo-parcel based crop identification by integrating high spatial-temporal resolution imagery from multi-source satellite data. Remote Sens. 9, 1298. https://doi.org/10.3390/rs9121298.

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J., Zhang, L., 2020. Deep learning in environmental remote sensing: achievements and challenges. Remote Sens. Environ. 241, 111716 https://doi.org/10.1016/j.rse.2020.111716.

Zanaga, D., Van De Kerchove, R., De Keersmaecker, W., Souverijns, N., Brockmann, C., Quast, R., Wevers, J., Grosu, A., Paccini, A., Vergnaud, S., Cartus, O., Santoro, M., Fritz, S., Georgieva, I., Lesiv, M., Carter, S., Herold, M., Linlin, L., Tsendbazar, N., Raimoino, F., Arino, O., 2021. ESA WorldCover 10 M 2020 v100 1–27. https://doi.org/10.5281/zenodo.5571936.

Zeiler, M.D., Fergus, R., 2014. Visualizing and understanding convolutional networks. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer, pp. 818–833. https://doi.org/10.1007/978-3-319-10590-1_53.

Zhai, H., Zhang, H., Zhang, L., Li, P., 2018. Cloud/shadow detection based on spectral indices for multi/hyperspectral optical remote sensing imagery. ISPRS J. Photogramm. Remote Sens. 144, 235–253. https://doi.org/10.1016/j.isprsjprs.2018.07.006.

Zhang, L., Weng, Q., 2016. Annual dynamics of impervious surface in the Pearl River Delta, China, from 1988 to 2013, using time series Landsat imagery. ISPRS J. Photogramm. Remote Sens. 113, 86–96. https://doi.org/10.1016/j.isprsjprs.2016.01.003.

Zhang, Y., Guindon, B., Cihlar, J., 2002. An image transform to characterize and compensate for spatial variations in thin cloud contamination of Landsat images. Remote Sens. Environ. 82, 173–187. https://doi.org/10.1016/S0034-4257(02)00034-2.

Zhang, Liangpei, Zhang, Lefei, Du, B., 2016. Deep learning for remote sensing data: a technical tutorial on the state of the art. IEEE Geosci. Remote Sens. Mag. 4, 22–40. https://doi.org/10.1109/MGRS.2016.2540798.

Zhang, L., Weng, Q., Shao, Z., 2017. An evaluation of monthly impervious surface dynamics by fusing Landsat and MODIS time series in the Pearl River Delta, China, from 2000 to 2015. Remote Sens. Environ. 201, 99–114. https://doi.org/10.1016/j.rse.2017.08.036.

Zhang, C., Sargent, I., Pan, X., Li, H., Gardiner, A., Hare, J., Atkinson, P.M., 2018. An object-based convolutional neural network (OCNN) for urban land use classification. Remote Sens. Environ. 216, 57–70. https://doi.org/10.1016/j.rse.2018.06.034.

Zhang, X., Liu, L., Chen, X., Gao, Y., Xie, S., Mi, J., 2021. GLC_FCS30: global land-cover product with fine classification system at 30 m using time-series Landsat imagery. Earth Syst. Sci. Data 13, 2753–2776. https://doi.org/10.5194/essd-13-2753-2021.

Zhu, Z., Woodcock, C.E., 2012. Object-based cloud and cloud shadow detection in Landsat imagery. Remote Sens. Environ. 118, 83–94. https://doi.org/10.1016/j.rse.2011.10.028.

Zhu, Z., Woodcock, C.E., 2014. Continuous change detection and classification of land cover using all available Landsat data. Remote Sens. Environ. 144, 152–171. https://doi.org/10.1016/j.rse.2014.01.011.

Zhu, X.X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: a comprehensive review and list of resources. IEEE Geosci. Remote Sens. Mag. 5, 8–36. https://doi.org/10.1109/MGRS.2017.2762307.

Zupanc, A., 2017. Improving Cloud Detection with Machine Learning. https://medium.com/sentinel-hub/improving-cloud-detection-with-machine-learning-c09dc5d7cf13.