



Characterization and mapping of photovoltaic solar power plants by Landsat imagery and random forest: A case study in Gansu Province, China

Xinxin Wang^a, Xiangming Xiao^b, Xi Zhang^a, Hui Ye^c, Jinwei Dong^d, Qiang He^a, Xubang Wang^e, Jianquan Liu^e, Bo Li^f, Jihua Wu^{e,*}

^a Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, National Observations and Research Station for Wetland Ecosystems of the Yangtze Estuary, Institute of Biodiversity Science and Institute of Eco-Chongming, School of Life Sciences, Fudan University, Shanghai, 200438, China

^b Department of Microbiology and Plant Biology, Center for Earth Observation and Modeling, University of Oklahoma, Norman, OK, 73019, USA

^c School of Tourism and Geography, Jiujiang University, Jiujiang, 332005, China

^d Key Laboratory of Land Surface Pattern and Simulation, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, 100101, China

^e State Key Laboratory of Herbage Improvement and Grassland Agro-Ecosystems, And College of Ecology, Lanzhou University, Lanzhou, 730000, China

^f Key Ministry of Education Key Laboratory for Transboundary Ecosystem Security of Southwest China and Yunnan Key Laboratory of Plant Reproductive Adaptation and Evolutionary Ecology and Centre for Invasion Biology, Institute of Biodiversity, School of Ecology and Environmental Science, Yunnan University, Kunming, Yunnan, 650504, China

ARTICLE INFO

Handling Editor: Zhifu Mi

Keywords:

Photovoltaic solar power plants
Landsat imagery
Random forest
Morphological characteristics
Land use changes
Gansu province

ABSTRACT

Numbers and sizes of photovoltaic solar power plants have grown unprecedentedly over the last few years in China, which aims to achieve a carbon emission peak by 2030 and carbon neutrality by 2060. Thus, timely and accurate monitoring of photovoltaic solar power plants is crucial to the design and management of renewable electricity systems in China. Random forest algorithm has been used to map photovoltaic solar power plants at multiple scales, however, it always causes several salt-and-pepper noises, limiting its application at larger spatial scales. Here we first develop a photovoltaic solar power plant mapping method through integrating time series Landsat imagery, random forest, and morphological characteristics. Then we apply this method in Gansu Province, which has abundant solar and wind energy resources and provide large amounts of potential lands for photovoltaic development, and generate the annual photovoltaic maps from 2015 to 2020. We further analyze the spatial-temporal dynamics of sizes and areas of photovoltaic solar power plants and major land cover conversion of expansive photovoltaic regions. Finally, we discuss the reliability, uncertainties, implications, and future development of our improved methods. We find our photovoltaic mapping method can remove most of salt-and-pepper noises effectively, and the resultant maps in Gansu for 2020 have very high accuracies with user's and producer's accuracies of 97.57% and 99.22%, respectively. There are 165.29 km² photovoltaic solar power plants in Gansu for 2020, and most of which are located in the northwestern Gansu. In addition, the photovoltaic with patch size > 1 km² and ≤ 2 km² (53.4 km², 32.3%) has largest patch number (39, 15.7%). The improved photovoltaic mapping methods and further analysis in this study provide critical information for accurate and automatic classification of photovoltaic solar power plants in the future, as well as the environmental and sustainable development of solar energy in China.

1. Introduction

Global energy demand is increasing to fulfill the needs of the growing human population as fossil fuel consumption has increased significantly over the past half-century, around eight-fold since 1950, and roughly doubling since 1980, leading to global warming and creating problems

related to climate change (Aryal et al., 2021; Ritchie et al., 2022). Photovoltaic (PV) technology is widely accepted as one of practical solutions to climate change and environmental pollution due to the burning of fossil fuels (Creutzig et al., 2017; Jiang et al., 2020), and is believed as one crucial approach to the achievement of the United Nations (UN) Sustainable Development Goals (SDGs) (Kruitwagen et al.,

* Corresponding author.

E-mail address: wjh@lzu.edu.cn (J. Wu).

<https://doi.org/10.1016/j.jclepro.2023.138015>

Received 16 March 2023; Received in revised form 19 June 2023; Accepted 5 July 2023

Available online 6 July 2023

0959-6526/© 2023 Published by Elsevier Ltd.

2021; Xia et al., 2022a). Globally, the capacity of PV solar power generation has grown by 41% per year since 2009, and increased to 423 GW (GW) at the end of 2018, among which 100 GW was newly installed in 2018, contributing 55% of new renewable energy capacity (Dunnnett et al., 2020; Kruitwagen et al., 2021). Furthermore, 4240 GW of solar PV is projected to be deployed by 2040 (Kruitwagen et al., 2021).

China is the world's largest carbon emission economy, and a high proportion of its electricity is still generated from fossil fuel combustion, which contributes to more than 40% of the national carbon emissions (Jiang et al., 2020; Wei et al., 2020). Since 2007, China has spent great efforts in developing the PV industry to transform its energy structure, and its total installed PV capacity increased from 100 MW in 2007 to 205,000 MW in 2019, with a compound annual growth rate of 79.8% (Dong et al., 2020). Synchronously, China's carbon emission intensity decreased by 48.1% compared with that in 2005, achieving the reduced carbon intensity target of 40–45% by 2020 that proposed during the 2009 Copenhagen Climate Change Conference (Chen et al., 2022). As China aims to achieve a carbon emission peak by 2030 and carbon neutrality by 2060, PV solar power generation is expected to keep growing rapidly across China (Lu et al., 2021; Mallapaty, 2020).

However, the PV solar cells might also have potential environmental and ecological impacts, such as, the changes in albedo and land use and land cover, which may give feedback to local climate changes (Stamford and Azapagic, 2018; Washington and Meehl, 1993), the habitat and biodiversity loss under the PV system during manufacturing (Hastik et al., 2015; Holland et al., 2019; Turney and Fthenakis, 2011). Thus, timely, automated, and accurate monitoring of PV solar power plants is crucial to the design, operation, and management of increasingly renewable electricity systems in China, as well as the assessment of social and ecological impacts (e.g., biodiversity, ecological functions) of large-scale PV deployment (Dunnnett et al., 2020; Xia et al., 2022a).

Fieldwork and bottom-up reporting are traditional methods for mapping and tracking PV solar power plants (Jiang et al., 2021). However, their unacceptable costs in terms of time and efforts and lack of geospatial information and precision restrict their widespread use at large spatial scales (Jiang et al., 2020, 2021). With the advances of spatial-temporal resolutions of sensors, satellite-based remote sensing provides images to track land cover changes in near real-time at multiple scales and resolutions (Hou et al., 2022; Huang et al., 2022; Tao et al., 2023; Wang et al., 2020a, 2020b; Yang et al., 2022). Visual interpretation has been widely used in previous studies for mapping PV solar power plants; however, it is often labor-intensive, time consuming, and difficult to be extended to large regions at non-acquisition times (Wang et al., 2020a, 2020b). As the increase in freely available time series remote sensing data, machine learning method (e.g., random forest) or objected-oriented analysis method have been the most popular methods for mapping and tracking PV solar power plants at the global (Dunnnett et al., 2020; Kruitwagen et al., 2021), national (Xia et al., 2022a; Zhang et al., 2022), or regional scales (Jiang et al., 2021; Tao et al., 2023).

Random Forest (RF) algorithm was developed by Breiman (2001), and has much more advantages than other machine learning classifiers, such as nonlinear mining capabilities, overfitting prevention, fast training, and quantitative description of the contribution of variables (Phalke et al., 2020; Zhou et al., 2020). However, in pixel-based RF classification, sudden disturbances in the image signal and different land covers with the same spectrum or the same land cover with a different spectrum can cause salt-and-pepper noises (Zhang et al., 2022). Currently, the common method to remove these noises is visual interpretation (Xia et al., 2022a, 2022b; Zhang et al., 2022), but it is too costly in time and effort to be applied at larger spatial scales. Thus, in light of the above difficulties, how to monitor PV solar power plants using random forests from time-series satellite images with fewer noises requires further research.

Gansu Province, located in the northwest of China, has abundant solar and wind energy resources, and is one of the earliest provinces to study and develop solar power plants in China. The installed PV capacity

increased to 5060 MW in 2014, ranking first in China (Tian and Xue, 2016). Furthermore, the desertified land area of Gansu Province is $\sim 192,100 \text{ km}^2$, accounting for 45.12% of the province's total land area and 18% of China's total desertified land area, therefore providing large amounts of potential land for PV development (Zhou and Li, 2022). Thus, it is an ideal region to develop high-resolution mapping algorithms of PV and analyze the spatial-temporal changes of PV solar power plants.

In this study, we select Gansu Province as study area to (1) develop a basic approach to identifying PV solar power plants based on time-series Landsat, random forest machine learning method, and the morphological characteristics of PV; (2) generate the detailed and accurate PV maps of Gansu Province from 2015 to 2020; (3) evaluate the accuracy of resultant PV maps; (4) analyze the spatial distribution and temporal dynamics of PV from 2015 to 2020 in Gansu; (5) investigate the major land cover conversions in those expansive PV regions; and (6) discuss the reliability, uncertainties, implications, and future development of PV mapping methods.

2. Materials and methods

2.1. Study area

The Gansu Province, which has 14 cities, is located at the intersection of three main plateaus of China (i.e., the Loess Plateau, Qinghai-Tibet Plateau, and Inner Mongolia Plateau), with a total area of $45.59 \times 10^4 \text{ km}^2$ (Wang et al., 2022a) and a wide range of elevations ranging from 600 to 5600 m above sea level (Liu et al., 2020) (Fig. 1a). The uneven precipitation, which decreases from southeast to northwest with annual means ranging from 40 to 800 mm, results in the desert-oasis landscape in the western Hexi Corridor due to dry climate, scarce precipitation, and strong evapotranspiration and complex landform types (dominated by mountains and hills) and rich forest and grassland resources in the eastern part (including the Gannan Plateau and part of the Loess Plateau) in Gansu (Wang et al., 2022a; Wen et al., 2017).

2.2. Datasets

2.2.1. Landsat data

We use time series Landsat surface reflectance (SR) datasets as satellite resources to map PV solar power plants in this study. Landsat acquires images at 16-day revisit cycle and 30-m spatial resolution, and all images have undergone necessary pre-processing in Google Earth Engine (GEE) cloud-based geospatial processing platform, including radiometric calibration and atmospheric correction (Wang et al., 2022b, 2023). We also use the quality assurance band that is generated by the CFMask algorithm to identify bad quality observations, including clouds and cloud shadows. As the images from September to December in a year usually have fewer clouds, cloud shadows, and less green vegetation cover, and have been used to map PV solar power plants at multiple scales (Xia et al., 2022a; Zhang et al., 2022), here we use all the available Landsat imagery from September to December in GEE of Gansu Province to identify PV solar power plants in 2015 and 2020. However, we find that Landsat imagery has very limited number of good-quality observations from September to December in 2020 (Fig. 1b) as over 17% pixels of Gansu have fewer than 4 good-quality observations and over 43% pixels have fewer than 8 good-quality observations, especially those pixels in southeastern Gansu. In order to avoid the effects of limited Landsat data on PV mapping, we integrate the Landsat images within 3-year time period to map the PV map of 2020 (2019–2021) and map of 2015 (2014–2016). This approach enables us to have enough numbers of good-quality observations in each 3-year period as over 93% pixels have more than 12 good-quality observations, and only 3.5% pixels have fewer than 4 good-quality observations (Fig. 1c). Such data compositing approach has been widely used to map global and regional land use and land covers, such as coastal wetlands (Wang et al., 2021)

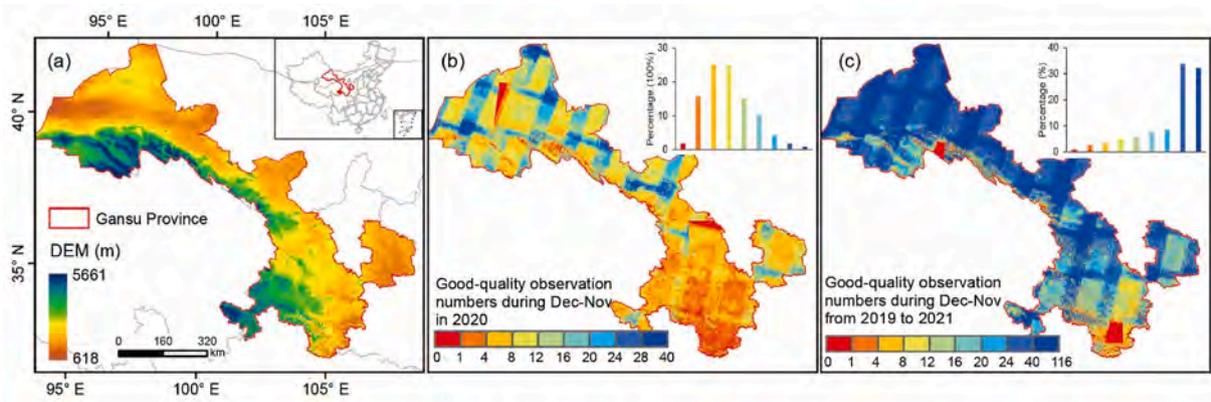


Fig. 1. Study area. (a) Location of Gansu Province in China and the distribution of digital elevation model (DEM) in Gansu; (b) Spatial distribution of good-quality observation numbers of Landsat imagery in 2020; (c) Spatial distribution of good-quality observation numbers of Landsat imagery during 2019–2021.

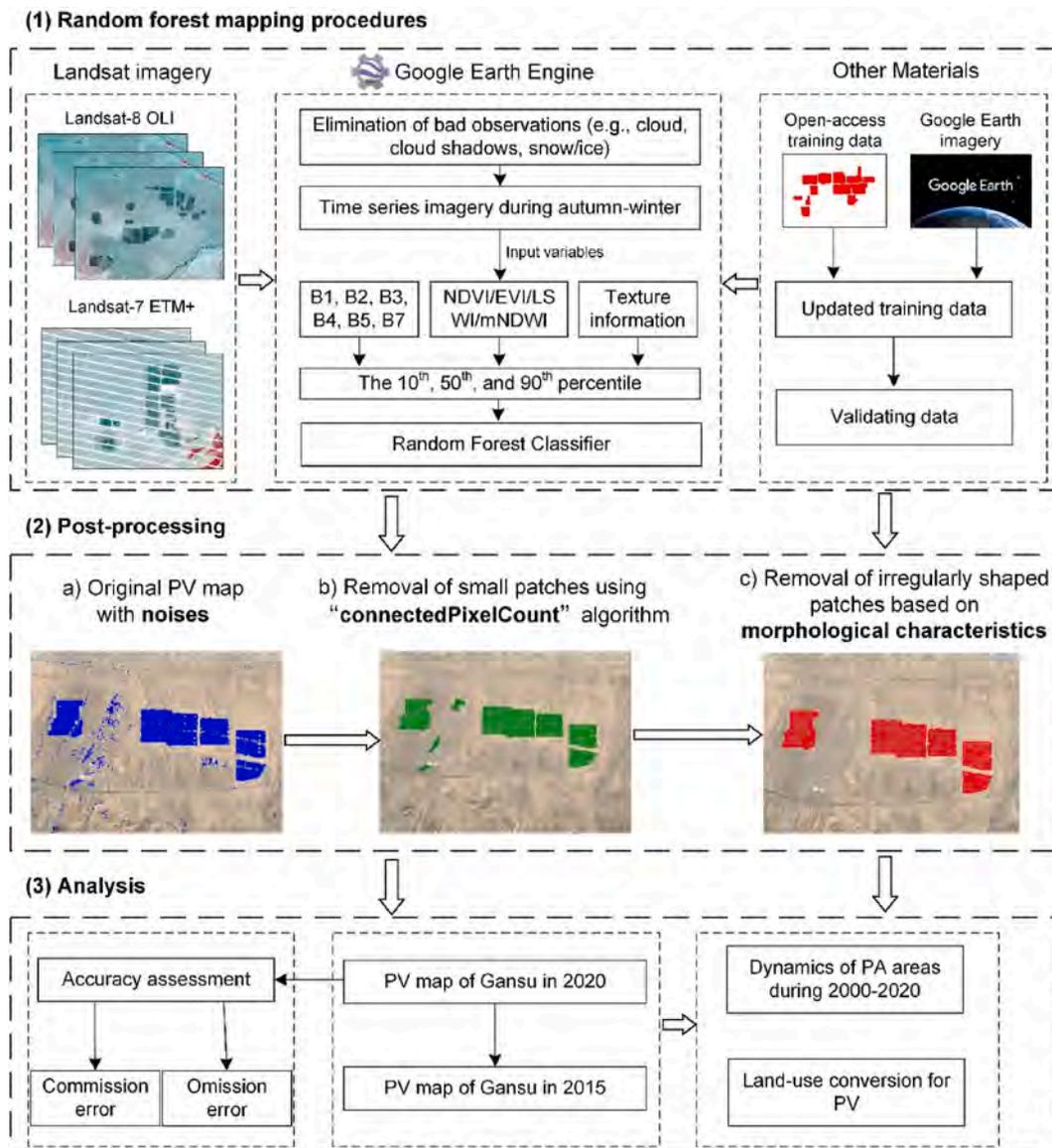


Fig. 2. Workflow of this study. It includes three parts: (1) generation of photovoltaic (PV) solar power plant maps using time series Landsat imagery, random forest algorithm, and Google Earth Engine (GEE) platform; (2) post-processing for removing noises based on patch areas and morphological characteristics; (3) accuracy assessment of resultant PV maps; and (4) further analyses, including spatial-temporal dynamics of PV and land use conversion for PV solar power plants.

and tidal flats (Murray et al. 2019, 2022).

2.2.2. Published PV maps

Zhang et al. (2022) published the PV power plant map in 2020 of China using Landsat in autumn season and random forest method, and released PV solar power plant map and the training dataset in 2020. Here we acquire them for inter-comparison and improvement of our training sample dataset.

2.2.3. Land use and land cover map in 2015

The 30-m land use and land cover (LULC) map of Gansu Province for 2015 (LULC2015) is acquired from the Data Center for Resources and Environmental Sciences, Chinese Academy of Sciences (<https://www.resdc.cn/DOI/DOI.aspx?DOIID=54>) to analyze the major land cover changes in those expansive PV regions during 2015–2020 (Fig. S1).

2.3. Methods

In this study, we use the pixel-based random forest (RF) algorithm to map the PV solar power plants in Gansu Province in the GEE platform for 2015 and 2020, and analyze their spatial-temporal dynamics. The workflow of this study is divided into four parts (Fig. 2): (1) generation of PV power plant maps by using time series Landsat imagery, RF algorithm, and GEE platform; (2) post-processing for removing noises based on patch size and morphological characteristics; (3) accuracy assessment of resultant PV maps, and (4) further analyses, including spatial-temporal dynamics of PV, and land use conversion for PV solar power plants.

2.3.1. Sample data for algorithm training and map accuracy assessment

The locations and numbers of algorithm-training samples are crucial for the accuracy and stability of RF classification (Zhou et al., 2020). Here we primarily collect published PV and Non-PV training point released by Zhang et al. (2022) based on the Landsat-OLI imagery during autumn season as our training samples. But we find that it does not cover all the PV solar power plant types in Gansu, especially in southeastern Gansu, where PV solar power plants are rarely labeled (Fig. 3a, j), and thus we further enrich the training dataset by manually selecting and labeling PV solar power plants to ensure that the samples can be evenly distributed in Gansu Province. Finally, 2142 PV sample points and 3013

non-PV sample points are collected in this study, and 70% of them are used to train the RF classifier and the rest of them are used to validate the final PV map.

2.3.2. Random forest classification

The first step for RF classification is to determine input variables. In this study, we collect three major kinds of input variables, including six original Landsat bands (B1-B6, and B7), four vegetation and water indices (NDVI, EVI, LSWI, and mNDWI, Eqs (1)–(4)), and four texture indices (homogeneity, correlation, contrast, and entropy). Homogeneity is a measurement of lack of variability or the amount of local similarity in the scene; correlation is a measure of grey level linear dependencies in the image, and high correlation values denote a linear relationship between the grey levels of pixel pairs; contrast is a measure of the amount of local variation in pixel values between neighboring pixels, and it is high for regions exhibiting large local variations and is the opposite of homogeneity; entropy is a measure of the degree of disorder in an image, and larger value occurs when the image is texturally non-uniform or heterogeneous (Franklin et al., 1996; Zhou et al., 2017). All these four texture indices can be calculated using the ee.Image.glmTexture algorithm provided by GEE team (<https://developers.google.com/earth-engine/apidocs/ee-image-glmtexture>). Then, we choose the maximum, medium, and minimum of each variable to train the RF classifier. In order to reduce the effects of unmasked cloud and poor-quality observations, we calculate the 90th percentile, 50th percentile, and 10th percentile of each variable from September to December within 3-year time period as the input variable of RF classifier.

The next step for the RF classifier is to set the number of trees (Ntree), which is the key parameter in the RF classifier as larger Ntree contributes to higher accuracy but longer run time and possible supersaturation. In this study, we test the sensitivity of different Ntree during 10–500, and find that the accuracy is kept stable when Ntree is greater than 350 (Fig. 4a). Thus, the number of Ntree is set to 350 in the RF classifier. Furthermore, we calculate the importance of each input variable (Fig. S2), and the results show that the original bands and spectral indices have greater importance than texture information. We also set the rest of the parameters of the RF classifier at GEE's default following Zhang's study (Zhang et al., 2022) for better inter-comparison.

$$NDVI = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}} \quad (1)$$

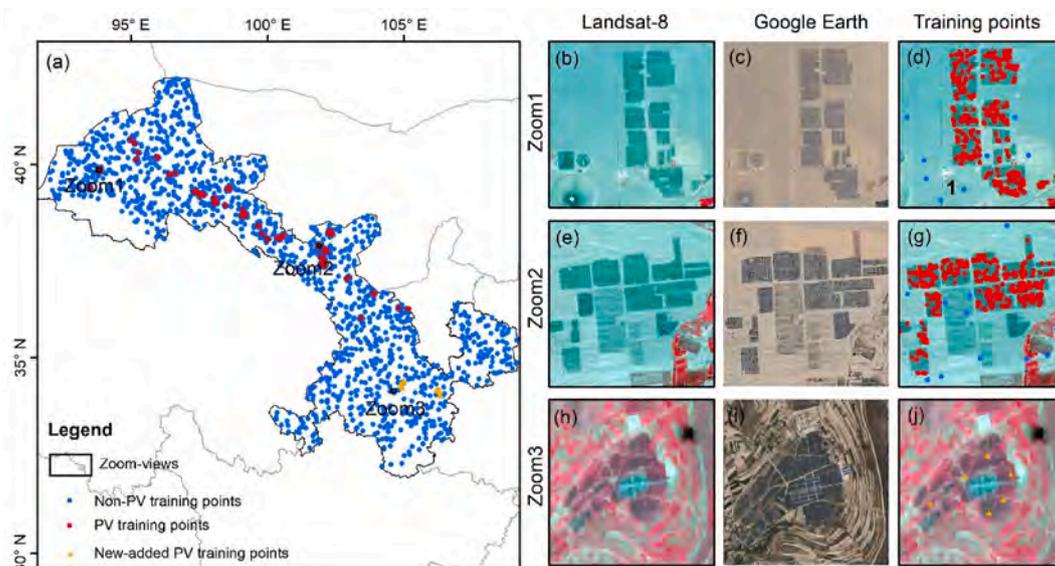


Fig. 3. Spatial distribution of training points of photovoltaic (PV) and non-PV solar power plants. (a) Spatial distribution of training points of PV and non-PV; (b–j) Images from Landsat-8, Google Earth, and Landsat with training points of PV in three Zoom-views. The Landsat-8 images are acquired in 2020-07-27 (Zoom1), 2020-08-18 (Zoom2), and 2020-07-19 (Zoom3), respectively.

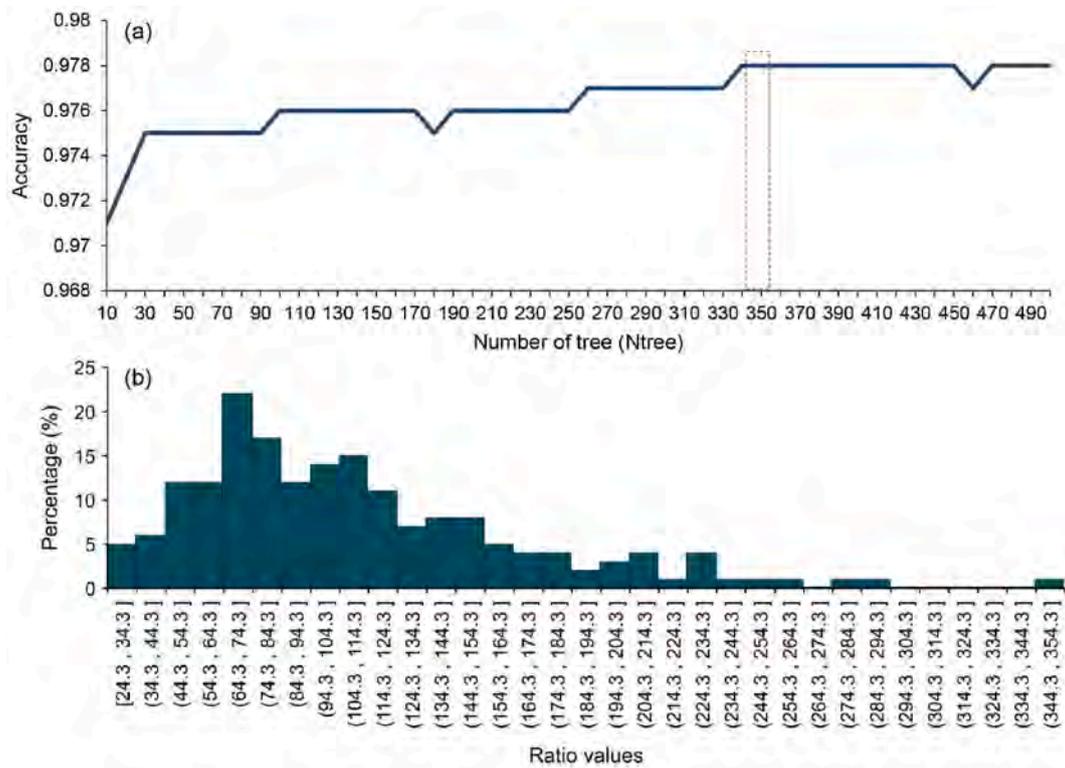


Fig. 4. Learning curve of random forest classifier (a) and the distribution of ratio (area/perimeter) values within PV polygons (b).

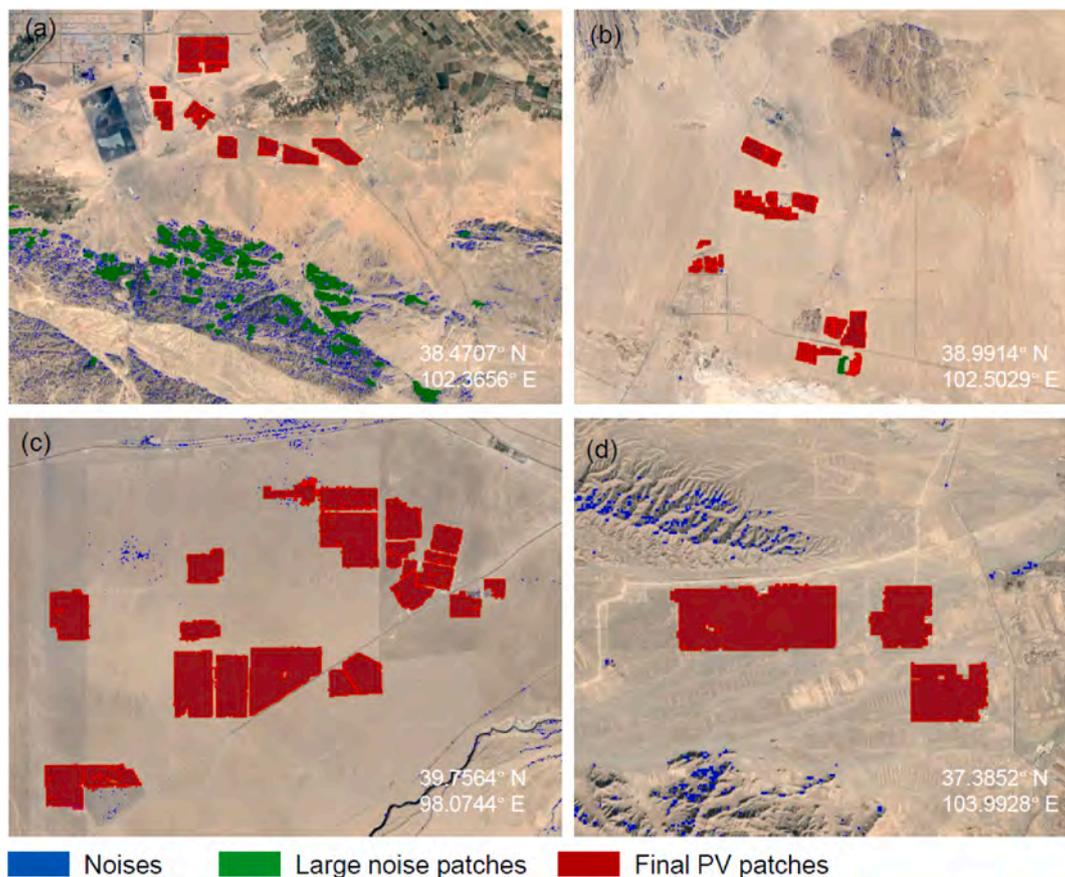


Fig. 5. Post processing for removing noises while mapping PV. (a) Preliminary PV maps with noises in 2020; (b–d) PV maps after removing small-size noises by using the “connectedPixelCount()” function in GEE platform.

$$EVI = 2.5 \times \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + 6 \times \rho_{red} - 7.5 \times \rho_{blue} + 1} \quad (2)$$

$$LSWI = \frac{\rho_{nir} - \rho_{swir}}{\rho_{green} + \rho_{swir}} \quad (3)$$

$$mNDWI = \frac{\rho_{green} - \rho_{swir}}{\rho_{green} + \rho_{swir}} \quad (4)$$

where ρ_{blue} , ρ_{green} , ρ_{red} , ρ_{nir} , and ρ_{swir} are the surface reflectance values of blue, green, red, near-infrared, and shortwave infrared bands in Landsat images.

2.3.3. Post-processing for removing noises based on morphological characteristics

By using the RF classification, we first generate the preliminary PV maps in 2015 and 2020 of Gansu Province. Then we filter continuous patches of PV that pixels continuously connect in at least one of eight directions to reduce the salt-and-pepper noise with small sizes by using the “connectedPixelCount()” function in GEE (Fig. 5). In addition, we also find that there are some large nose patches with irregular characteristics after removing small patches, here we calculate the ratio of patch area to patch perimeter to represent the morphological characteristics (Eq. (5)), and explore its frequency distribution of all the PV training samples (Fig. 4b). The results show that almost all the PV samples have ratio > 24, and thus we use the criteria of ratio ≤ 24 to remove those remaining irregular noise patches. Then, we generate the final PV maps of Gansu Province in 2015 and 2020.

$$ratio = \frac{Area}{Perimeter} \quad (5)$$

2.4. Accuracy assessment of the resultant PV maps

Stratified random sampling approach, along with very-high spatial resolution images from Google Earth Pro, is the most widely used and robust approach in accuracy assessment of land cover classification (Murray et al., 2019; Pekel et al., 2016; Wang et al., 2020b), by which we validate the PV map in 2020. In order to calculate the user’s accuracy (measure of commission error) of the resultant map in this study, for each 0.05 by 0.05 grid cell, two points are generated randomly within the final map of Gansu Province in 2020 acquired using the above-mentioned improved algorithms, and a total of 288 random points are finally selected. Each point is checked and interpreted visually in Google Earth imagery determining its land cover types (PV or non-PV). We use the 30% of training dataset (643 points) introduced in the Section 2.3.1 to calculate the producers’ accuracy (measure of omission error) (Table 1).

2.5. Inter-comparison of PV solar power plant maps

We also acquire the published PV map in the same year from Zhang et al. (2022) for inter-comparison. They released the PV power plant map in 2020 of China using Landsat in autumn season and random forest method. Thus, we compare the PV solar power plan areas from our study with Zhang’s results.

Table 1
Accuracy assessment results of PV map for 2020.

| | Classified | Misclassified | Total | Accuracy (100%) |
|---------------------|------------|---------------|-------|-----------------|
| User’s accuracy | 281 | 7 | 288 | 97.57 |
| Producer’s accuracy | 638 | 5 | 643 | 99.22 |

2.6. Area calculation of different land covers from the expansion of PV solar power plants

We first extract the 30-m land cover pixels for 2015 within the PV map in Gansu Province and different cities. Then we count the numbers of 30-m land cover pixels and calculate the areas (pixel number × 900 m²) and percentages of different land cover types in Gansu and different cities.

3. Results

3.1. Accuracy assessment of the annual PV map of Gansu in 2020

The user’s accuracy (measure of commission error) for the PV map in this study is 97.57%, and the producer’s accuracy (measure of omission error) for the PV map is 99.22% (Table 1). The producer’s accuracy is smaller than the user’s accuracy because we integrate the 3-year Landsat images during September and December, which enables us to use many more Landsat image for classification and have much smaller omission PV polygons. The results indicate that the resultant PV maps generated in this study have a good agreement between mapped pixels and ground-referenced pixels.

3.2. Patch numbers and sizes of PV solar power plants

PV solar power plants are unevenly distributed across Gansu Province in 2020. There are 248 patch numbers and 16,529 ha patch areas of PV solar power plants for 2020 in Gansu, and the PV with patch size > 1 km² and ≤ 2 km² (40, 15.7%) has largest patch areas (53.4 km², 32.3%) (Fig. 6), followed by those with patch size > 2 km² and ≤ 3 km² (20.7 km², 12.5%). However, the PV solar power plants with patch size > 0.1 km² and ≤ 0.2 km² has largest patch number (44, 17.7%) (Fig. 6a). Furthermore, most of PV solar power plants are located in the north-western Gansu. From the heat map, four larger PV density regions are found in our study, including western Jiuquan, Jiayuguan, Jinchang, and Tianshui (Fig. 7a). Statistical analysis for each city of Gansu shows

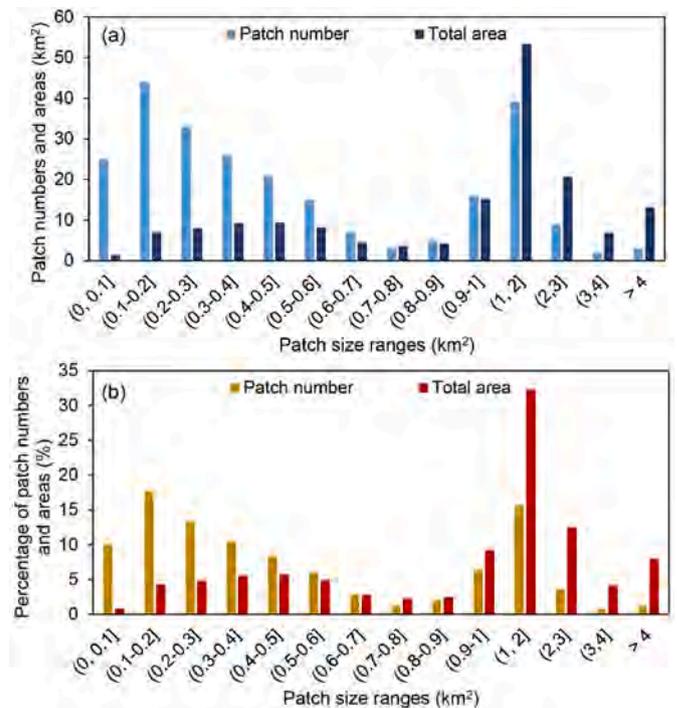


Fig. 6. PV solar power plants for 2020 in Gansu Province. (a) Patch numbers and areas (km²) of PV for 2020; (b) Percentages of patch numbers and areas of PV for 2020.

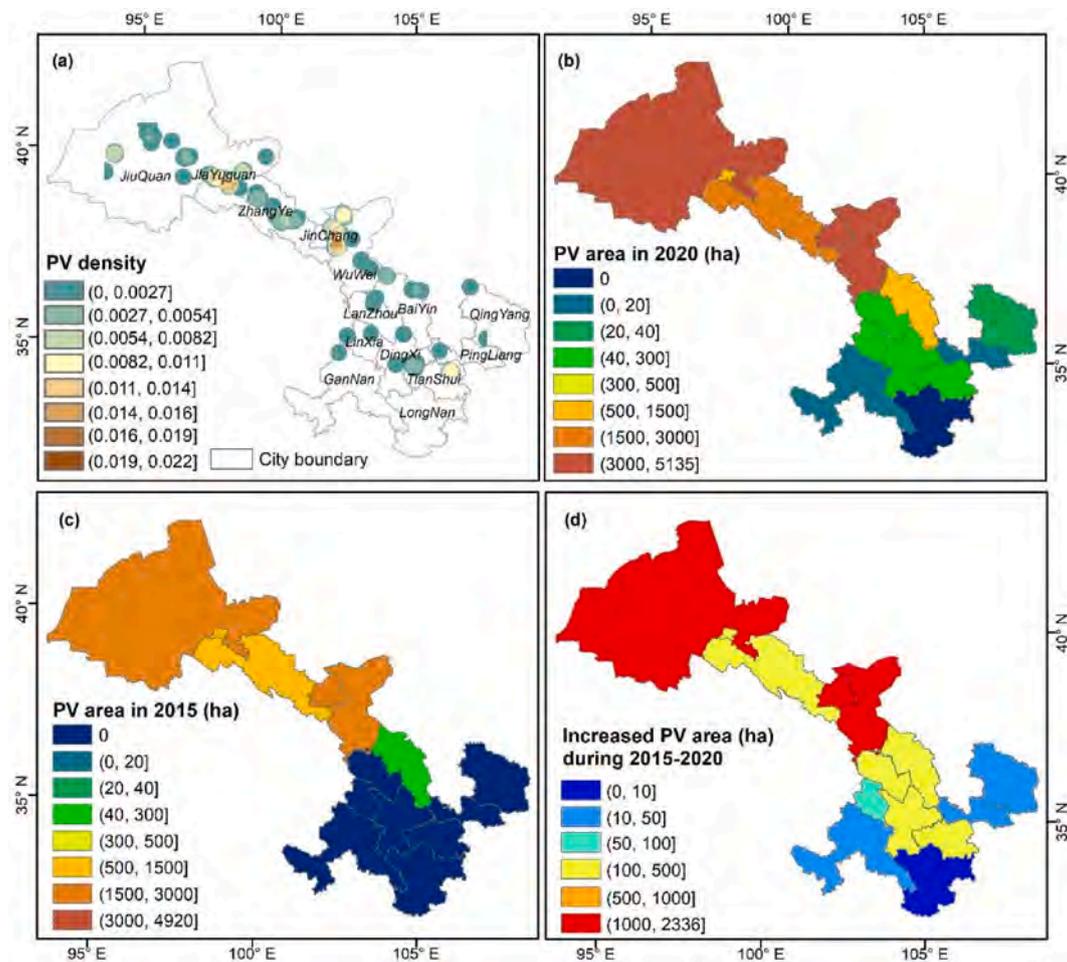


Fig. 7. PV areas (ha) for 2015 and 2020 in Gansu. (a) PV density in 2020; (b) PV area in 2020; (c) PV area in 2015; and (d) increased PV areas from 2015 to 2020.

that Jiuquan has the largest PV area (5135 ha) in 2020, followed by Jinchang (3682 ha), Wuwei (3032 ha), Zhangye (1796 ha), and Jiayuguan (1423 ha) (Fig. 7b). Longnan in the southeastern Gansu has no PV solar power plants (0 ha), and Gannan (10 ha) and Pingliang (11 ha) also have much smaller PV areas than other cities.

3.3. Temporal dynamics of PV areas in Gansu during 2015–2020

PV solar power plants also have divergent increased areas by city in Gansu during 2015–2020. The spatial distribution of zoom-views demonstrates the detailed expansion of PV (Fig. 8). Only 9929 ha PV are found by 2015 in Gansu, most of which are located in northwestern Gansu (Fig. 7c), indicating about 6600 ha new PV solar power plants are constructed during 2015–2020. Furthermore, Jiuquan (2335 ha), Jinchang (1102 ha), and Wuwei (1088 ha) have much larger increased area than other cities, and the increased PV areas in the three cities account for 69% of the total area in Gansu (Fig. 7d).

3.4. Land cover change from the expansion of PV solar power plants

The land-use change analysis shows that the newly constructed PV solar power plants in Gansu are mainly converted from four land cover types: gobi (63.9%), sparse grasslands (12.7%), other built-up lands (e. g., large industrial areas, and mines) (8.9%), and croplands (7.6%) (Fig. 9a). Different cities in Gansu also have divergent conversion proportions of major land-use types (Fig. 9b). All the gobi is found to be converted to PV in four cities: Jiuquan (89.5%), Wuwei (76.8%), Jinchang (74.6%), Zhangye (48.9%), and Jiayuguan (45.5%). The

conversion from sparse grasslands is found in almost all the cities, especially in Lanzhou (64.7%), Baiyin (43.0%), Jiayuguan (39.3%), and Zhengye (32.5%). Croplands have the largest conversion proportions in Dingxi (84.6%) and Linxia (70.8%). Other built-up lands are mainly found in four cities (Jinchang, Zhangye, Jiayuguan, and Wuwei) with large PV areas. Baiyin has relatively larger conversion from salina (26.6%) than other cities.

4. Discussion

4.1. Annual maps of PV solar power plants in Gansu Province

Timely and accurate monitoring of PV solar power plants is crucial to design, operation, and management of increasingly renewable electricity systems and assessment of the social and ecological impacts of large-scale PV deployment (Dunnett et al., 2020; Xia et al. 2022a, 2022b). Random forest algorithm has been widely used to map different land covers at multiple scales, such as global coastal wetlands (Murray et al. 2019, 2022), and national PV maps (Xia et al., 2022a; Zhang et al., 2022). However, it always causes several salt-and-pepper noises induced by sudden disturbances in the image signal and different land covers with the same spectrum or the same land cover with a different spectrum, costing too much time and effort to remove them and limiting its application at larger spatial scales (Zhang et al., 2022). In this study, we find the PV solar power plants usually have regular characteristics and large nose patches have irregular characteristics, thus, we calculate the ratio of patch area to patch perimeter for representing the morphological characteristics and removing those irregular noises automatically

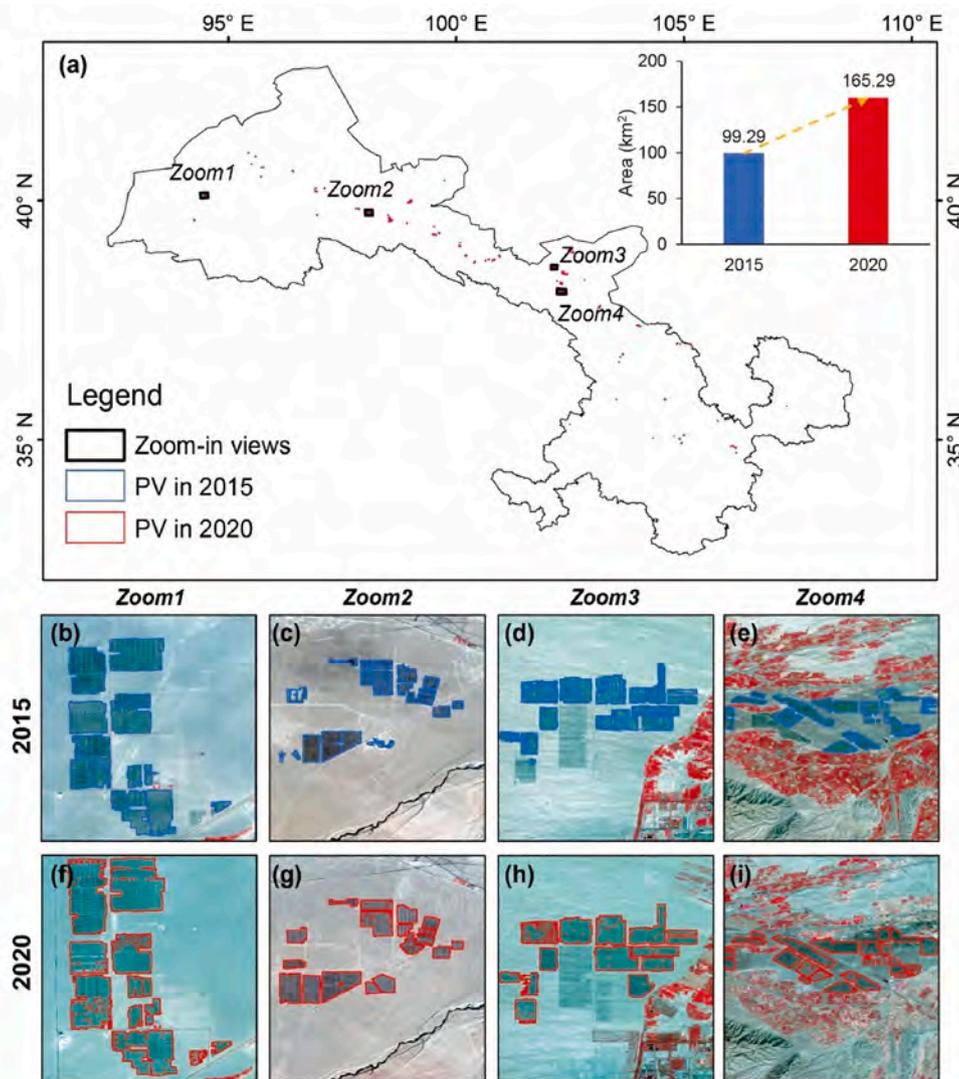


Fig. 8. Expansion of PV during 2015–2020. (a) Spatial distribution of PV in 2015 and 2020, as well as four zoom-in views and increased PV areas in Gansu during 2015–2020; (b–e) Detailed information of PV in 2015 in four zoom-in views; (f–i) Detailed information of PV in 2020 in four zoom-in views.

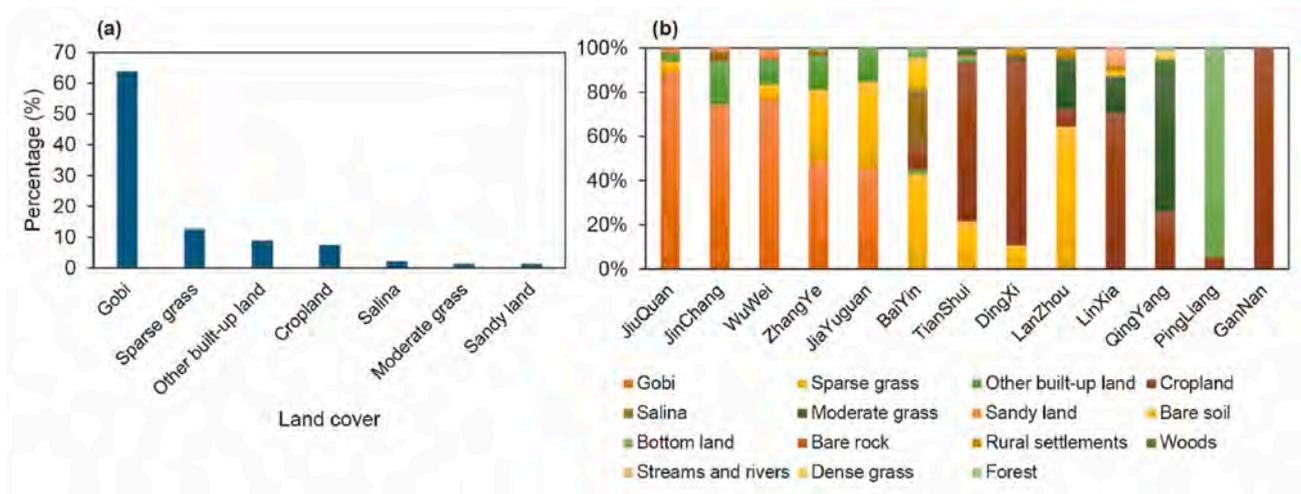


Fig. 9. Land cover of expansive PV during 2015–2020 in Gansu Province. (a) Percentage of different land covers within expansive PV solar power plants in Gansu; (b) Percentage of different land covers within expansive PV solar power plants in different cities of Gansu.

and effectively. In addition, we integrate the Landsat images after September within 3-year time period to map the PV maps, which enables us to have enough numbers of good-quality observations and smaller omission errors of our final PV maps. Furthermore, we also enrich the training dataset in Gansu Province based on Zhang's dataset (Zhang et al., 2022). Thus, the accurate and automatic classification method of PV solar power plants, and much more Landsat imagery and training samples contribute to the smaller commission error of our resultant PV maps in Gansu Province.

The comparison between our PV map of Gansu Province and other studies also shows the good accuracy of our results. Our study reports 16,529 ha PV areas, and Zhang's study reports 17,855 ha PV in 2020 (Zhang et al., 2022). The different input Landsat imagery, training dataset, and different input variables of RF classifier might explain the inconsistency between Zhang's map and our study (Table 2). First, in terms of input image data, Zhang's study composited the Landsat-8 images of autumn 2020 (September to November) as Landsat resources and used the whole year of 2020 (January to December) substitute in the regions where the imagery quality of autumn 2020 was poor. However, we use all the Landsat imagery (including Landsat-7 and Landsat-8) from September to December during 2019–2021 as the satellite resources to generate final PV map. Second, in terms of the training samples, we enrich and improve the training samples based on Zhang's dataset to make sure they are evenly covered across Gansu Province (Fig. 3). Third, in terms of input variables, Zhang's study used nine variables from the Landsat-8 SR images data, including six original bands and three spectral indices, as input variables, but our study collects six original Landsat bands (B1–B6, and B7), four vegetation and water indices, and four texture indices in the RF classifier. The larger numbers of Landsat imagery, training samples, and input variables can include more PV solar power plants and more accurate PV information in our study (Fig. 10c and d). Furthermore, our study uses the morphological characteristics to remove those irregular noises automatically, contributing to the much more effective PV mapping in our study (Table 2). However, as the PV polygons from Zhang's study included several non-PV pixels around PV polygons (Fig. 10a and b), resulting in the overestimation of PV areas in their study.

Table 2
Comparison of PV solar power plant maps between our study and Zhang's study.

| | Our study | Zhang's study | Impact of differences between two studies on the resultant PV maps |
|----------------------------------|---|--|---|
| Methods for removing noises | Ratio of patch area to patch perimeter to represent the morphological characteristics | Visual interpretation | Much more automatically and effectively in our study for removing noises |
| Landsat dataset | All the Landsat imagery from September to December during 2019–2021 | Landsat-8 images of autumn 2020 | Much larger number of good-quality Landsat observations contributes to more PV information and accurate PV boundaries |
| Training dataset | 2142 PV samples points and 3013 non-PV sample points | 2061 PV samples points and 3013 non-PV sample points | Larger training points of PV in our study contribute to PV results with greater accuracy |
| Input variables of RF classifier | Six original Landsat bands, four vegetation and water indices, and four texture indices | Six original bands and three spectral indices | More input variables in our study contribute to PV results with greater accuracy |

4.2. Sources of errors in annual PV solar power plant maps

It has been a challenging task to generate annual maps of PV solar power plants at large spatial domains. Accuracy of annual PV maps is affected by several factors: (1) the mixture of PV and occupied land, (2) land cover types with similar spectral features with PV, (3) enough training and validation samples representing all the PV types, and (4) noises caused by RF classifier and limited satellite image quality.

Complex landscapes of PV and its occupied land cover types, as well as those types with similar spectral features with PV pose great challenges for PV mapping. For example, some plastic-covered sheds, factory, regularly shaped factories, and biological soil crust may be misclassified as PV in the RF classification (Fig. S3). The numbers of occupied land cover types can be much larger when PV solar power plants are mapped at much larger spatial scales (Xia et al., 2022a), and thus enough training and validation samples representing all the PV types are of great importance to the high accuracy of final maps. Furthermore, some bad-quality observations (e.g., clouds, cloud shadows) may remain after quality filtering, which is likely to generate some noises in the final maps (Zhang et al., 2022). Fortunately, these noises usually have irregular shape and cover only a very small proportion, and are removed through morphological characteristics and visual interpretation in our study.

There are still some omission errors in the PV mapping algorithms. For Landsat images at 30-m spatial resolution, those PV panels with areas smaller than 30m by 30m cannot be identified and mapped. In addition, some PV solar power plants, which have lower density in mountainous areas and have non-PV land cover within a PV polygon, tend to be misclassified as non-PV objects (Figs. S3 and S4). For these reasons, the areas of PV in our study are likely to be underestimated.

4.3. Implications and future development of PV mapping

The RF algorithm integrating Landsat imagery and morphological characteristics for mapping PV solar power plants at 30-m spatial resolution is critical for better understanding of the detailed spatial information of PV solar power plants, and can provide invaluable information for the design, operation, and management of increasingly renewable electricity systems. This mapping strategy can be used to (1) monitor the distribution and trajectory of PV at national or global scales; and (2) track the dynamics of PV over the last decades. However, this method may still have some errors when it is used in the tropical regions because of more frequent cloud cover and in the mountainous areas with considerable terrain shadows. In the future, we can integrate more optical satellite data at similar spatial resolutions with Landsat data, such as Sentinel-2 and Worldview 3, as well as the microwave images from synthetic aperture radar (SAR), to enrich the good-quality observation numbers and generate PV maps with higher accuracy.

In addition, the expansion of PV solar power plants can have potential environmental and ecological impacts (Stamford and Azapagic, 2018; Washington and Meehl, 1993). The most direct impact of PV development on the environment is the changes of albedo resulting from the land use changes, which can give feedback to local climate changes (Washington and Meehl, 1993). The manufacturing of PV solar cells also involves different kinds of hazardous materials during the extraction of solar cells, posing substantial threats to the environment at much larger spatial scales (Alami et al., 2020). Furthermore, the PV systems also have great influence on terrestrial biodiversity and ecosystem functions (Holland et al., 2019) through the land-use changes under the PV system (Hastik et al., 2015) and installation activities (e.g. vegetation clearing, removal of upper soil layers) (Turney and Fthenakis, 2011). Therefore, it is necessary to develop technologies and methodologies to better monitor and understand interactions between environment, wildlife, ecosystems, and large-scale solar systems, as well as their potential feedbacks to climate. It is also recommended to develop new and high-efficiency technologies for recycling the PV waste to reduce the

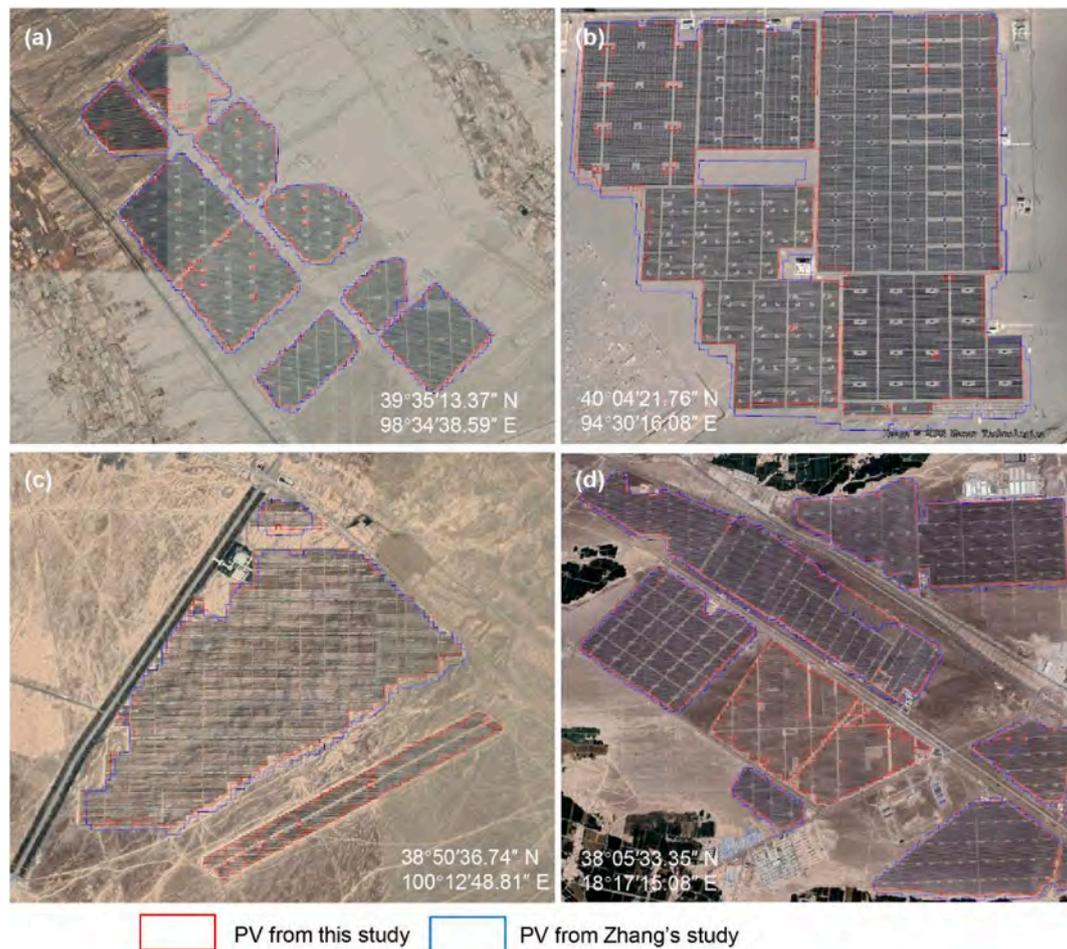


Fig. 10. Improvement for mapping PV in this study comparing with Zhang's study. (a–b) The PV map from Zhang's study with several non-PV pixels. (c–d) Larger PV power plant area in our study comparing with that from Zhang's study.

environmental risks and sustaining the raw materials supplies in the future.

5. Conclusions

Timely and accurate monitoring of PV solar power plants is crucial to design, operation, and management of increasingly renewable electricity systems and assessment of the social and ecological impacts of large-scale PV deployment. Random forest algorithm has been widely used to map PV solar power plants at multiple scales, but it always causes several salt-and-pepper noises, limiting its application at larger spatial scales. In this study, we calculate the ratio of patch area to patch perimeter to represent the morphological characteristics, and remove those noises automatically and effectively. For Gansu Province as example, we generate the annual PV solar power plant maps through integrating the Landsat images after September within 3-year time period in 2015 and 2020. We find our PV mapping method can remove most of salt-and-pepper noises effectively, and the resultant maps in Gansu for 2020 have very high accuracies with user's and producer's accuracies of 97.57% and 99.22%, respectively. There are 16,529 ha PV in Gansu for 2020, and most of which are located in the northwestern Gansu. Jiuquan, Jinchang, and Wuwei have much larger increased areas than other cities during 2015–2020, and the increased PV areas in the three cities account for 69% of the total increased area in Gansu. The newly constructed PV solar power plants in Gansu are mainly converted from gobi. The improved PV mapping methods in this study provide critical approaches for accurate and automatic classification of PV solar power plants at larger scales, and the resultant maps can greatly enhance

our understanding of the spatial distribution and temporal dynamics of PV power development, and the major land cover changes within the expansive PV regions in Gansu.

CRediT authorship contribution statement

Xinxin Wang: Conceptualization, Data curation, Formal analysis, Validation, Visualization, Writing – original draft, Writing – review & editing. **Xiangming Xiao:** Formal analysis, Writing – review & editing. **Xi Zhang:** Methodology, Software, Validation. **Hui Ye:** Methodology, Software, Validation. **Jinwei Dong:** Formal analysis, Writing – review & editing. **Qiang He:** Formal analysis, Writing – review & editing. **Xubang Wang:** Methodology, Software, Validation. **Jianquan Liu:** Formal analysis, Writing – review & editing. **Bo Li:** Formal analysis, Writing – review & editing. **Jihua Wu:** Funding acquisition, Supervision, Conceptualization, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that can have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgements

This work was supported by the Gansu Provincial Key Program of Science Fund (22JR5RA396), and Fundamental Research Funds for the Central Universities (lzujbky-2021-sp51).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jclepro.2023.138015>.

References

- Alami, A.H., Faraj, M., Aokal, K., Abu Hawili, A., Tawalbeh, M., Zhang, D., 2020. Investigating various permutations of copper iodide/FeCu tandem materials as electrodes for dye-sensitized solar cells with a natural dye. *Nanomaterials* 10, 784.
- Aryal, N., Ottosen, L.D.M., Kofoed, M.V.W., Pant, D., 2021. *Emerging Technologies and Biological Systems for Biogas Upgrading*. Denmark: Academic Press, Aarhus.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.
- Chen, J.D., Xu, C., Gao, M., Li, D., 2022. Carbon peak and its mitigation implications for China in the post-pandemic era. *Sci. Rep.* 12, 3473.
- Creutzig, F., Agoston, P., Goldschmidt, J.C., Luderer, G., Nemet, G., Pietzcker, R.C., 2017. The underestimated potential of solar energy to mitigate climate change. *Nat. Energy* 2, 1–9.
- Dong, H.Q., Zeng, B., Wang, Y.Q., Liu, Y.X., Zeng, M., 2020. China's solar subsidy policy: government funding yields to open markets. *IEEE Power Energy Mag.* 18, 49–60.
- Dunnnett, S., Sorichetta, A., Taylor, G., Eigenbrod, F., 2020. Harmonised global datasets of wind and solar farm locations and power. *Sci. Data* 7, 130.
- Franklin, S.E., Wulder, M.A., Lavigne, M.B., 1996. Automated derivation of geographic window sizes for use in remote sensing digital image texture analysis. *Comput. Geosci.* 22, 665–673.
- Hastik, R., Basso, S., Geitner, C., Haida, C., Poljanec, A., Portaccio, A., Vrscaj, B., Walzer, C., 2015. Renewable energies and ecosystem service impacts. *Renewable Sustainable Energy Rev.* 48, 608–623.
- Holland, R.A., Scott, K., Agnolucci, P., Rapti, C., Eigenbrod, F., Taylor, G., 2019. The influence of the global electric power system on terrestrial biodiversity. *Proc. Natl. Acad. Sci. U. S. A* 116, 26078–26084.
- Hou, T., Sun, W., Chen, C., Yang, G., Meng, X., Peng, J., 2022. Marine floating raft aquaculture extraction of hyperspectral remote sensing images based decision tree algorithm. *Int. J. Appl. Earth Obs. Geoinf.* 111, 102846.
- Huang, K., Yang, G., Yuan, Y., Sun, W., Meng, X., Ge, Y., 2022. Optical and SAR images Combined Mangrove Index based on multi-feature fusion. *Science of Remote Sensing* 5, 100040.
- Jiang, H., Yao, L., Lu, N., Qin, J., Liu, T., Liu, Y.J., Zhou, C.H., 2021. Multi-resolution dataset for photovoltaic panel segmentation from satellite and aerial imagery. *Earth Syst. Sci. Data* 13, 5389–5401.
- Jiang, M.K., Li, J.S., Wei, W.D., Miao, J.W., Zhang, P.F., Qian, H.Q., Liu, J.M., Yan, J.Y., 2020. Using existing infrastructure to realize low-cost and flexible photovoltaic power generation in areas with high-power demand in China. *iScience* 23, 101867.
- Kruitwagen, L., Story, K.T., Friedrich, J., Byers, L., Skillman, S., Hepburn, C., 2021. A global inventory of photovoltaic solar energy generating units. *Nature* 598, 604–610.
- Liu, L.J., Liang, Y.J., Hashimoto, S., 2020. Integrated assessment of land-use/coverage changes and their impacts on ecosystem services in Gansu Province, northwest China: implications for sustainable development goals. *Sustain. Sci.* 15, 297–314.
- Lu, X., Chen, S., Nielsen, C.P., Zhang, C.Y., Li, J.C., Xu, H., Wu, Y., Wang, S.X., Song, F., Wei, C., He, K.B., McElroy, M.B., Hao, J.M., 2021. Combined solar power and storage as cost-competitive and grid-compatible supply for China's future carbon-neutral electricity system. *Proc. Natl. Acad. Sci. U. S. A* 118, e2103471118.
- Mallapaty, S., 2020. How China could be carbon neutral by mid-century. *Nature* 586, 482–483.
- Murray, N.J., Phinn, S.R., DeWitt, M., Ferrari, R., Johnston, R., Lyons, M.B., Clinton, N., Thau, D., Fuller, R.A., 2019. The global distribution and trajectory of tidal flats. *Nature* 565, 222–225.
- Murray, N.J., Worthington Thomas, A., Bunting, P., Duce, S., Hagger, V., Lovelock Catherine, E., Lucas, R., Saunders Megan, I., Sheaves, M., Spalding, M., Waltham Nathan, J., Lyons Mitchell, B., 2022. High-resolution mapping of losses and gains of Earth's tidal wetlands. *Science* 376, 744–749.
- Pekel, J.F., Cottam, A., Gorelick, N., Belward, A.S., 2016. High-resolution mapping of global surface water and its long-term changes. *Nature* 540, 418–422.
- Phalke, A.R., Ozdogan, M., Thenkabail, P.S., Erickson, T., Gorelick, N., Yadav, K., Congalton, R.G., 2020. Mapping croplands of Europe, Middle East, Russia, and central Asia using Landsat, random forest, and Google Earth engine. *ISPRS J. Photogrammetry Remote Sens.* 167, 104–122.
- Ritchie, H., Rosado, P., Roser, M., 2022. *Energy*. In: Our World in Data.
- Stamford, L., Azapagic, A., 2018. Environmental impacts of photovoltaics: the effects of technological improvements and transfer of manufacturing from Europe to China. *Energy Technol.* 6, 1148–1160.
- Tao, S., Rogan, J., Ye, S., Geron, N., 2023. Mapping photovoltaic power stations and assessing their environmental impacts from multi-sensor datasets in Massachusetts, United States. *Remote Sens. Appl.: Society and Environment* 30, 100937.
- Tian, M., Xue, H., 2016. Problems and countermeasures of new energy development in Gansu Province, China. *Gansu Social Sciences* 6, 208–213.
- Turney, D., Fthenakis, V., 2011. Environmental impacts from the installation and operation of large-scale solar power plants. *Renewable Sustainable Energy Rev.* 15, 3261–3270.
- Wang, L.G., Zhu, R., Yin, Z.L., Chen, Z.X., Fang, C.S., Lu, R., Zhou, J.Q., Feng, Y.L., 2022a. Impacts of land-use change on the spatio-temporal patterns of terrestrial ecosystem carbon storage in the Gansu province, northwest China. *Rem. Sens.* 14, 3164.
- Wang, X., Xiao, X., Qin, Y., Dong, J., Wu, J., Li, B., 2022b. Improved maps of surface water bodies, large dams, reservoirs, and lakes in China. *Earth Syst. Sci. Data* 14, 3757–3771.
- Wang, X., Xiao, X., Xu, X., Zou, Z., Chen, B., Qin, Y., Zhang, X., Dong, J., Liu, D., Pan, L., Li, B., 2021. Rebound in China's coastal wetlands following conservation and restoration. *Nat. Sustain.* 4, 1076–1083.
- Wang, X., Xiao, X., Zhang, X., Wu, J., Li, B., 2023. Rapid and large changes in coastal wetland structure in China's four major river deltas. *Global Change Biol.* 29, 2286–2300.
- Wang, X., Xiao, X., Zou, Z., Chen, B., Ma, J., Dong, J., Doughty, R.B., Zhong, Q., Qin, Y., Dai, S., Li, X., Zhao, B., Li, B., 2020a. Tracking annual changes of coastal tidal flats in China during 1986–2016 through analyses of Landsat images with Google Earth Engine. *Rem. Sens. Environ.* 238, 110987.
- Wang, X., Xiao, X., Zou, Z., Hou, L., Qin, Y., Dong, J., Doughty, R.B., Chen, B., Zhang, X., Chen, Y., Ma, J., Zhao, B., Li, B., 2020b. Mapping coastal wetlands of China using time series Landsat images in 2018 and Google Earth Engine. *ISPRS J. Photogrammetry Remote Sens.* 163, 312–326.
- Washington, W.M., Meehl, G.A., 1993. Greenhouse sensitivity experiments with penetrative cumulus convection and tropical cirrus albedo effects. *Clim. Dynam.* 8, 211–223.
- Wei, W.D., Hao, S.J., Yao, M.T., Chen, W., Wang, S.S., Wang, Z.Y., Wang, Y., Zhang, P.F., 2020. Unbalanced economic benefits and the electricity-related carbon emissions embodied in China's interprovincial trade. *J. Environ. Manag.* 263, 110390.
- Wen, X.H., Wu, X.Q., Gao, M., 2017. Spatiotemporal variability of temperature and precipitation in Gansu province (northwest China) during 1951–2015. *Atmos. Res.* 197, 132–149.
- Xia, Z., Li, Y., Guo, X., Chen, R., 2022a. High-resolution mapping of water photovoltaic development in China through satellite imagery. *Int. J. Appl. Earth Obs. Geoinf.* 107, 102707.
- Xia, Z.L., Li, Y.J., Chen, R.S., Sengupta, D., Guo, X.N., Xiong, B., Niu, Y.L., 2022b. Mapping the rapid development of photovoltaic power stations in northwestern China using remote sensing. *Energy Rep.* 8, 4117–4127.
- Yang, G., Huang, K., Sun, W., Meng, X., Mao, D., Ge, Y., 2022. Enhanced mangrove vegetation index based on hyperspectral images for mapping mangrove. *ISPRS J. Photogrammetry Remote Sens.* 189, 236–254.
- Zhang, X.H., Xu, M., Wang, S.J., Huang, Y.K., Xie, Z.Y., 2022. Mapping photovoltaic power plants in China using Landsat, random forest, and Google Earth Engine. *Earth Syst. Sci. Data* 14, 3743–3755.
- Zhou, B., Okin, G.S., Zhang, J.Z., 2020. Leveraging Google Earth Engine (GEE) and machine learning algorithms to incorporate in situ measurement from different times for rangelands monitoring. *Rem. Sens. Environ.* 236, 111521.
- Zhou, J., Li, S., 2022. PV and wind power generation boost economic development of Gansu Province and achieve "emission peak and carbon neutrality". *Sol. Energy* 4, 13–19.
- Zhou, J., Yan Guo, R., Sun, M., Di, T.T., Wang, S., Zhai, J., Zhao, Z., 2017. The Effects of GLCM parameters on LAI estimation using texture values from Quickbird Satellite Imagery. *Sci. Rep.* 7, 7366.