Extracting Solar Photovoltaic Installations from Satellite Images in China

Jinyue Wang [1,2],Jing Liu [1,2,*], Longhui Li [1,2]

[1] Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China.
[2] Key Laboratory of Virtual Geographic Environment (Nanjing Normal University), Ministry of Education, Nanjing 210023, China

* Correspondence: jingliugeo@njnu.edu.cn

Received: 25 October 2022; Revised: 10 March 2023; Accepted: 21 March 2023; Published 31 March 2023

Abstract: Developing renewable energy and achieving the energy transition is crucial to achieve the Sustainable Development Goals and carbon neutrality. Solar photovoltaic (PV) are rapidly expanding in China as a popular renewable energy technology. Remote sensing (RS) technology can effectively detect existing PV installations. Previous studies have focused on PV inspection on a small area scale. This study aims to effectively detect large-scale installations of PV in China using open multi-source remote sensing data. We used multi-source satellite images including Sentinel-1 SAR and Sentinel-2 multispectral data to construct classification features. We developed a random forest classifier on the Google Earth Engine platform to detect PV installations in China. Manually collected samples and existing PV database were used to assess the accuracy of the detection. The results show that we can detect PV devices nationwide with good accuracy (OA=98.90%, kappa=0.86). Our detection rate reaches 80.11% when compared with the previous database. The finding in this study can provide reference value for the future development and monitoring of solar renewable energy to support the Sustainable Development Goals.

Keywords: Google Earth Engine; Photovoltaic; Random Forest; Remote Sensing; Sentinel

1.0 Introduction

Global warming, fossil energy shortages and increasing environmental pollution are among the major challenges facing human society today. The United Nations has set up Sustainable Development Goals. The development of renewable energy sources plays a key role in energy transition to achieve sustainable development goals and carbon neutrality (Peng, Shuainan, Yanru, Shuchao, & Yuhu, 2021). Solar photovoltaic (PV) is developing rapidly as a popular renewable energy technology in various countries. China has also experienced a rapid expansion of solar PV installations. In 2017, China's total solar power generation accounted for one-third of global power generation (Li & Huang, 2020). In order to effectively estimate the potential of PV power generation and evaluate its impact on the ecological environment, accurately identification of PV installation is very important. For instance, detecting the spatial distribution of PV modules can provide a reference for better operation and planning of power systems (Erdener, Feng, Doubleday, Florita, & Hodge, 2022). Some of the openly shared PV data has also been referenced several times in urban construction and landscape planning (Wu & Biljecki, 2021).

Compared with traditional methods such as household surveys and interconnection filings, satellite remote sensing has the advantage to detect PV in terms of completeness, spatial and temporal resolution. Currently, there are several PV products at national scale in China, including the Global Power Plant Database (GPPD) (Byers et al., 2019), the global inventory of photovoltaic solar energy generating units (Kruitwagen et al., 2021), and the China's photovoltaic power plants in 2020(Zhang, Xu, Wang, Huang, & Xie, 2022). First of all, the GPPD released by the World Resources Institute is a comprehensive, open-source database of power plants around the world. The global inventory of PV units by Kruitwagen et al. is generated from Sentinel-2 and SPOT 6/7 satellite images using deep learning models (Kruitwagen et al., 2021). However, the SPOT 6/7 images are not freely available for users from China, thus limiting their application. Zhang et al. combined machine learning and visual interpretation methods with satellite data to map China's PV power plants in 2020 from Landsat 8 surface reflectance imagery, with a total area of 2917 square kilometers (Zhang et al., 2022). This database has the spatial resolution of 30m, which is lower than the 10m resolution from Sentinel-2 images. Due to lower spatial resolution, there may be more misclassification especially for small PV installations.

Therefore, the objective of this study is to detect PV installation in China using multi-source satellite images. We adopted the freely available Sentinel-1 SAR, Sentinel-2 multispectral data, and performed our analysis in China using the random forest classifier on the google earth engine platform. Manually collected samples and existing PV database were used to assess the classification results.

2.0 Study Area

We selected China as our study area, because China ranks among the world's leading countries in terms of PV's cumulative installed capacity and new installed capacity. China is located in eastern Asia and on the west coast of the Pacific Ocean. The territory is vast, with a total land area of 9.6 million square kilometers. The terrain shows a step-like distribution with high mountains in the west and low plains in the east. In China's vast and rich land, there are very rich solar light resources. More than 2/3 of the total area of the country has more than 2,000 hours of sunshine per year, and the annual radiation is more than 5000 MJ/m2. The distribution of solar energy in China is mainly higher in the northwest region than the southwest region.

3.0 Materials and Methods

3.1 Satellite Imagery

The satellite images we employed in this study are the Sentinel-2 multispectral and Sentinel-1 SAR data. Sentinel-2 is an optical satellite developed by the European Space Agency (ESA). Its multispectral sensor can take images in 13 wavelength bands, covering the visible, near infrared (NIR) and shortwave infrared (SWIR) spectral ranges (Drusch et al., 2012). We used the Sentinel-2 L2A surface reflectance product published on GEE, which has been atmospherically and geometrically corrected. We selected all Sentinel-2 surface reflectance images with cloud percentages below 10% between January and October 2020, resulting in 29765 images.

Sentinel-1, also consisting of two satellites, carries a C-band Synthetic Aperture Radar (SAR) that provides continuous all-weather observation images (Torres et al., 2012). In this study we used the Sentinel-1 Ground Range Detected (GRD) products (VV/VH) from January to March 2020 (spring), which have a resolution of 10m, and we acquired 4630 images of China on the GEE.

3.2 Training and test samples for PV detection

The study areas were classified into PV and non-PV (NPV) classes. First, when selecting PV samples, we randomly generated 4000 PV random points across the country based on the 2020 China Photovoltaic Power Plant Map Dataset released by Zhang et.al (Zhang et al.,

 E-ISSN: 2948-5231

2022). For these random points, a 10m*10m rectangular buffer was created for each point. In addition, we visually checked each sample with the help of very high-resolution satellite images available on the Google Earth Pro platform. Finally, 3,995 polygons were retained as PV samples. For NPV, we generated 6800 random points based on the ESA land cover product CCI-LC2020 (Pengyu et al., 2022), and built a 40m*40m rectangular buffers to filter NPV samples. Finally, we retained 6233 polygons as NPV samples. The spatial distribution of PV and NPV samples is shown in Figure 1. We use stratified random sampling method to divide the PV and NPV, where 70% is used as training sample 30% is used as test sample.

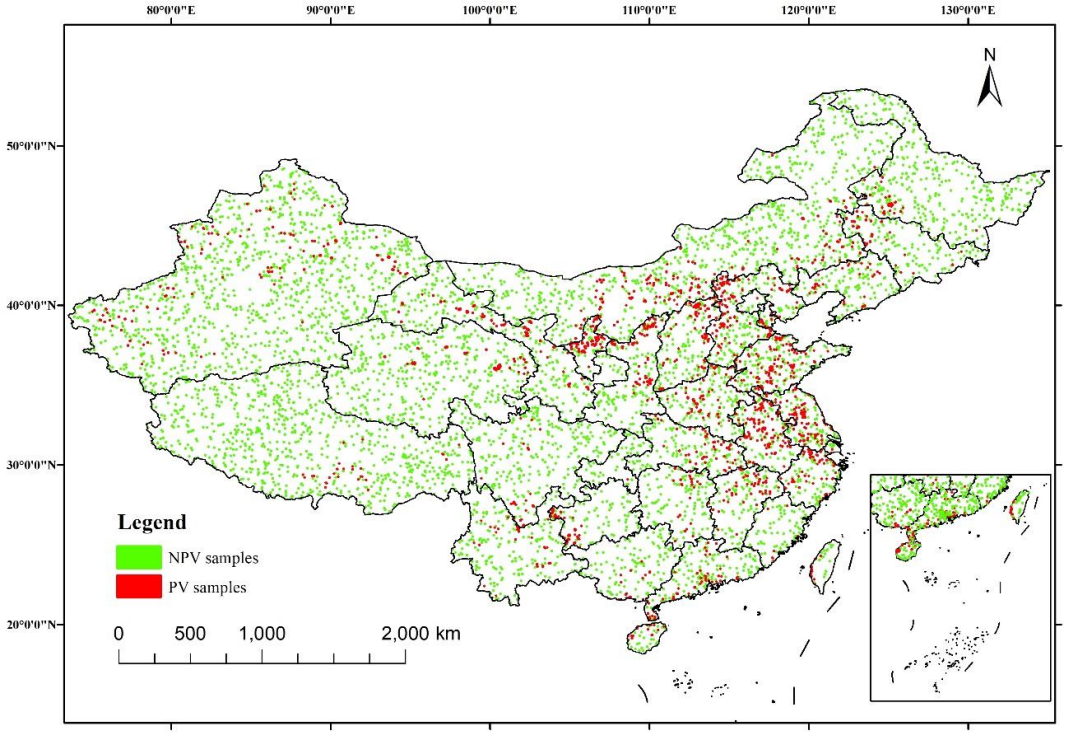

Figure 1: The distribution of PV and NPV samples.

3.3 Satellite Image Processing

Firstly, pre-processing operations such as cropping, time filtering, and cloud removal were performed on the Sentinel 1/2 images. We selected the administrative boundary of the study area to crop the original image, defined the cloud removal function and used the QA60 band mark to achieve cloud removal. After this, using the method of median filtering, the images of the whole year of 2020 were composited into one image.

We derived 13 variables from the Sentinel-2 images data, including ten original bands and three calculated indexes. The ten original bands included the B2, B3, B4, B5, B6, B7, B8, B8A, B11, and B12. The three indices included the normalized difference built-up index (NDBI) (Zha, Gao, & Ni, 2003), the normalized difference vegetation index (NDVI) (Tucker, 1979), and the modified normalized difference water index (mNDWI) (Xu, 2006) to enhance the signature of built-up, vegetation and water surface areas. In addition, the VV and VH bands of Sentinel-1 images were selected.

In addition, texture features of the above 15 variables were also calculated. In order to receive sunlight in a larger area, the layout of PV power plants is highly directional, and the texture features displayed on the image are obvious (Wei, Junxing, Guangzeng, Liang, & Deyong, 2021). Thereby, for each variable we calculated a total of 17 textural metrics, with the help of the ee.Image.glcmTexture function on GEE (Conners, Trivedi, & Harlow, 1984; Haralick, Shanmugam, & Dinstein, 1973). As a result, there are a total of 270 features used for this experiment. They are arranged in 2 groups listed in Table 1. With such a large number of features, we performed feature selection before classification to mitigate the effects of the Hughes phenomenon. We used the python scikit-learn library to calculate the importance of each feature in each group, the features with high importance scores were kept for subsequent classification.

Table 1: All 270 features under assessment in this study

Feature Groups	Number	Specific Features
S2 reflectance& Spectral indices &S1 polarization	15	B2, B3, B4, B5, B6, B7, B8, B8A, B11, B12, NDBI, NDVI, mNDWI ,VV, VH
Texture	255	Texture of B2, B3, B4, B5, B6, B7, B8, B8A, B11, B12, NDBI, NDVI, mNDWI, VV and VH

Note: S2 denotes Sentinel-2, while S1 denotes Sentinel-1

After feature selection, the random forest (RF) classifier was used to classify the images into PV and non-PV classes. Given the complex shape and spectral characteristics of PV installations, we chose to use the built-in random forest (RF) classifier in GEE for the importance evaluation of each feature component. As a widely used machine learning algorithm, RF has strong model generalization ability and good classification performance in the face of high level of data dimensionality and collinearity (Menghao, Guoqing, & Zhuangzhuang, 2022). The training and test samples for the RF classifier were as described in Section 2.3.

E-ISSN: 2948-5231

After classification, we employed the validation sample obtained in Section 2.3 to evaluate the results. The performance of the model with respect to the validation set is evaluated by comparing the confusion matrix of PV and NPV in the validation set using coefficient of kappa, overall accuracy, producer accuracy, and user accuracy. The kappa coefficient is widely used to test the consistency and evaluate the model performance. The producer accuracy indicates the proportion of true samples that are correctly judged as the target class, and the user accuracy indicates the proportion of samples judged as the target class in the classification graph.

4.0 Results

4.1 Feature importance

After feature selection, the original 270 classification features were eventually reduced to 14, which were added to the random forest classifier for the final detection of PV devices (Wang et al. (in review). These include five spectral reflectance features (B2, B12, B3, B11, B4), eight texture features (VV savg, VH_savg, NDBI_savg, slope_savg, MNDWI_savg, slope_imcorr1, B12_shade), and NDBI index features. According to the importance results from the RF classifier in GEE, it was found that the spectral reflectance of B2 had the highest impact on PV detection, followed by VV_sag and B12 in the second and third places (see Figure 2). We also found that the 'sum average' textures of NDBI, VV and VH all played a more important role in PV detection. The combination of spectral and textural features is therefore a more accurate way to detect PV installations.

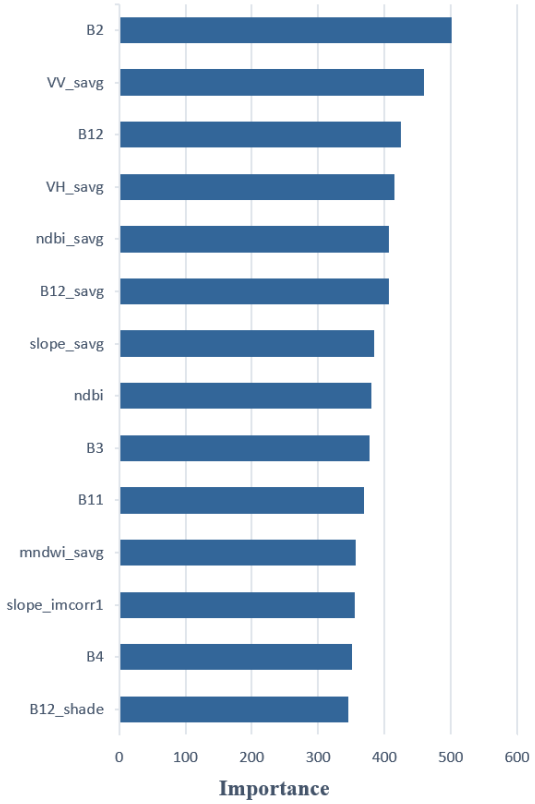

Figure 2: The importance of the 14 selected features in the RF classification

4.2 Classification Results

We added the selected classification features to the RF classifier to complete the detection of PV across China. The extraction results of some typical PV types are shown in Fig 3. For comparison, we used the blue and yellow colour to represent the PVs detected in the previous database (Zhang et al., 2022) and (Kruitwagen et al., 2021) respectively. The comparison reveals that the method proposed in this study can detect most of the PV. In some cases, our results had better delineation of PV than the previous database (see Figure 3(j)).

Further quantitative evaluation of the classification accuracy is shown in Table 2. The overall accuracy (OA) reached 98.45% and the kappa was 0.86, which demonstrates good performance of the method proposed in our study for PV detection in China.

Table 2: Accuracy of classification scheme

OA	Kappa	UA_PV	PA_PV	UA_NPV	PA_NPV
98.45%	0.86	98.85%	99.54%	95.79%	97.88%

E-ISSN: 2948-5231

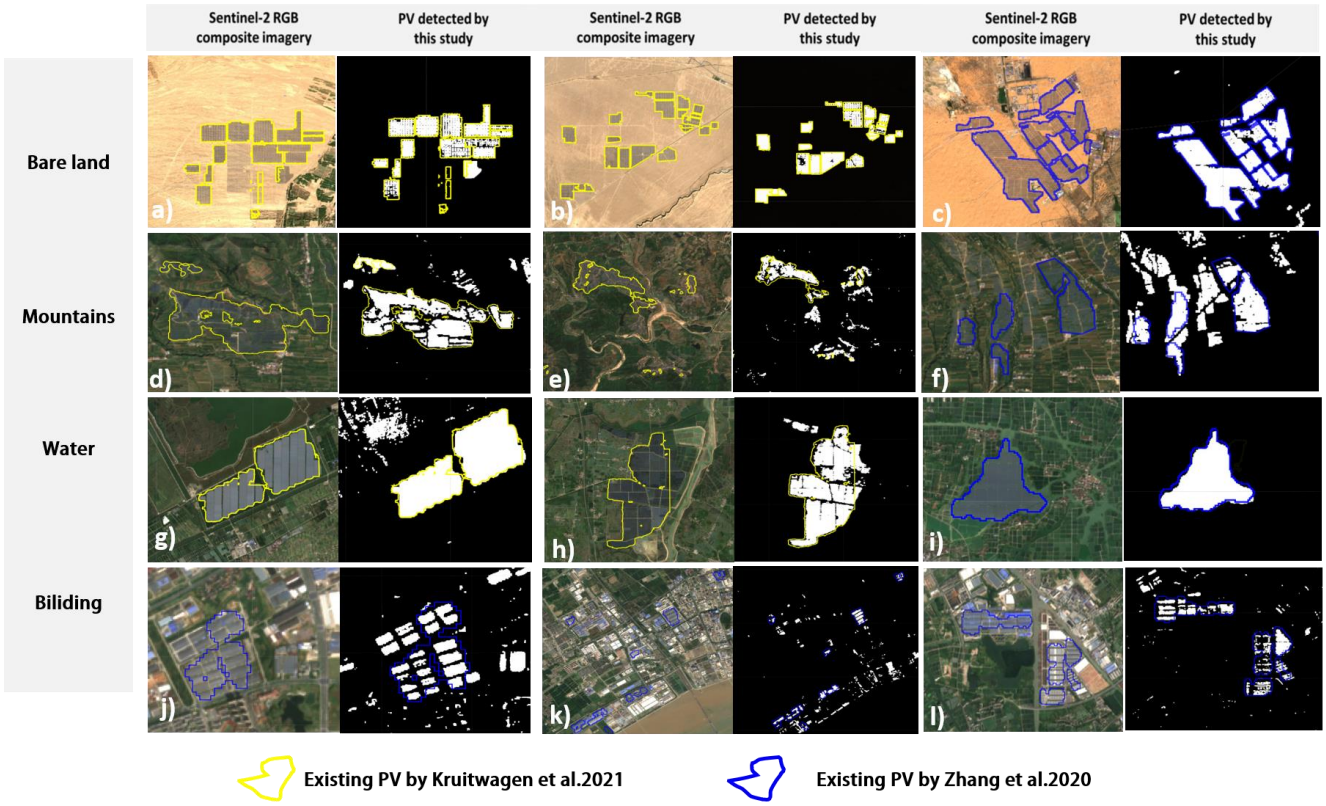

Figure 3. Classification results of PV stations in different land cover types. (a,b&c) Bare land; (d,e&f) Mountains and arable land; (g,h&i) Water (WPV); (j,k&l) Builing(roofspace)

4.3 Consistency evaluation with existing PV database

As PV installations are a relatively rare species among all land cover types, different perspectives have to be considered when performing the validation of the results. In addition to the classification accuracy, we also used existing PV database to evaluate the detection consistency. Due to the calculation capability limit of GEE, we selected the PV data of the Yangtze River Delta (YRD) region for consistency evaluation. We downloaded the PV installations in the YRD region from the previous Kruitwagen's dataset (Kruitwagen et al., 2021) from https://zenodo.org/record/5005868#.YzJjZXZByUI. A quantitative comparison of the PV detected in this study in YRD with the previous Kruitwagen's database is shown in Table 3. From Kruitwagen's dataset, we acquired 915 PV installations in the YRD region at "A" confidence rating. For each of these polygons, we calculated the proportion of detected PV area from this study. PV polygons with the proportion greater than 40% were considered as successful detection. The number of PVs correctly extracted was 732, and the detection rate reached 80.11%. In general, the PV installations extracted in this study were similar to those of the Kruitwagen's database.

Table 3: Consistency of our detection with existing PV database (In Yangtze River Delta Region)

Reference Database	Number of PV	TP	Detection Rate
Kruitwagen et al. 2021	915	732	80.11%

4.4 Limitation and future work

PV plants are composed of PV panels and the land they occupy. Across the vast tertiary of China, we found many features with similar spectral characteristics to PV, which led to misclassification which requires future refinement. Up till now, the main PV mapping errors in this study include bare rocks, mountain shadows and polyethylene on urban rooftops, as shown in Figure 4. These errors led to over-detection of PV and subsequent over-estimation of PV power output. Our recent work focus on the automatic elimination of these errors by image morphological operations. Future studies can explore using hyperspectral images as well as point-of-interest data to refine the PV detection.

In addition to the phenomenon of misclassification, there are also omission errors in the RF classification results. Some PV plants with low PV panel coverage can be incorrectly classified as non-PV objects. In addition, we found that the spectral mixing between solar PV panels and dense vegetation during the vegetation growing season also caused some mis-detection errors (Wang et al. (in review). In some cases, a change of land use led to omission errors. This illustrates the requirement for open sharing of nationwide PV datasets to ensure timeliness and accuracy in the context of China's rapidly growing PV industry.

As the industrial base and highly populated area, the power load in the YRD region is very high, while the land resources are very tight. This has led to the emergence of various PV types such as agricultural PV, fishery-PV and rooftop power stations. In the YRD region, our method detected many fishery-PV objects in ponds, lakes, and reservoirs. The development of these new PV industries contributes to the new energy transition and plays an important role in the sustainable development of the cities.

E-ISSN: 2948-5231

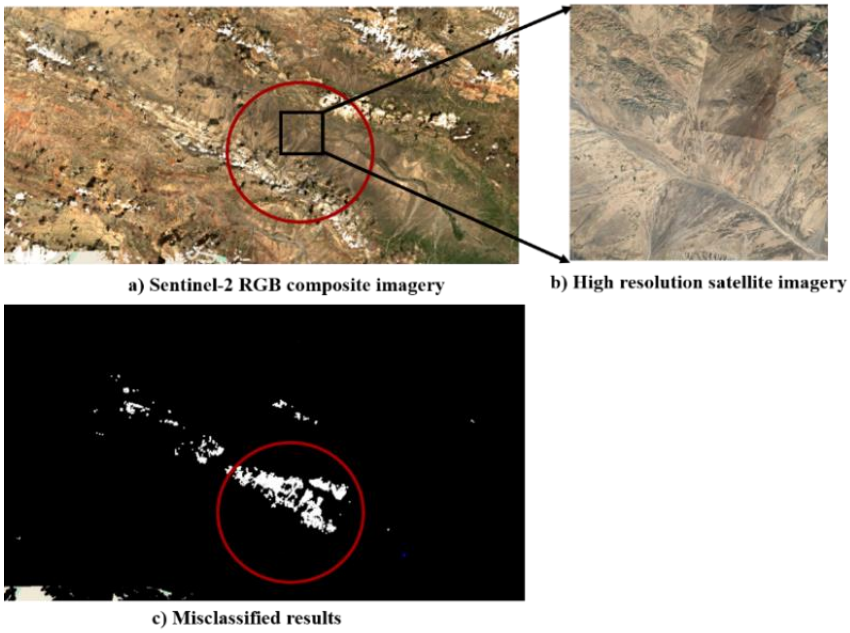

Figure 4: Bare rocks and mountain shadows misclassified as PV

5.0 Conclusions

In the context of the Sustainable Development Goals and carbon neutrality, solar PV is rapidly developing as a renewable energy technology worldwide. China's PV industry is expanding rapidly, and free medium-resolution satellite images can play a very important role in detecting the spatial distribution of PV. In this study, multi-source remote sensing data such as Sentinel-1 SAR and Sentinel-2 multispectral data were used to construct classification features and a random forest classifier on the Google Earth Engine platform were developed for detecting PV installations in China. The results showed that we can detect PV installations with good accuracy (OA = 98.90%, kappa = 0.86) at national scale. We achieved a detection rate of 80.11% when comparing to previous PV database. The results of this study can provide reference value for the future development and detection of the PV industry in China. From this study, we found that the main misclassifications include bare rocks, mountain shadows and roofing polyethylene. How to accurately and automatically eliminate these PV errors needs further research in the future

Acknowledgement: This work was supported by the Jiangsu Natural Science Foundation under Grant [BK20200722]; the Natural Science Foundation of the Jiangsu Higher Education Institutions of China under Grant [20KJB420001]; and the Jiangsu Province Innovation and Entrepreneurship Doctor Program.

Conflicts of Interest: The authors declare no conflict of interest.

References

Byers, L., Friedrich, J., Hennig, R., Kressig, A., Li, X., McCormick, C., & Valeri, L. M. (2019). A global database of power plants. World Resources Institute, Washington, DC. available at: https://www.wri.org/publication/global-database-power-plants (last access: 26 August 2020).

Conners, R. W., Trivedi, M. M., & Harlow, C. A. (1984). Segmentation of a high-resolution urban scene using texture operators. *Computer Vision, Graphics, and Image Processing, 25*(3), 273-310. https://doi.org/10.1016/0734-189X(84)90197-X

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., . . . Bargellini, P. (2012). Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment, 120*, 25-36. https://doi.org/10.1016/j.rse.2011.11.026

Erdener, B. C., Feng, C., Doubleday, K., Florita, A., & Hodge, B.-M. (2022). A review of behind-the-meter solar forecasting. *Renewable & Sustainable Energy Reviews, 160*. https://doi.org/10.1016/j.rser.2022.112224

Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics, 6*, 610-621.

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*(7882), 604-610. https://doi.org/10.1038/s41586-021-03957-7

Li, J., & Huang, J. (2020). The expansion of China's solar energy: Challenges and policy options. *Renewable and Sustainable Energy Reviews, 132*. https://doi.org/10.1016/j.rser.2020.110002

Menghao, Z., Guoqing, L., & Zhuangzhuang, P. (2022). Remote sensing crop classification method based on feature selection. *Science of Surveying and Mapping, 47*(3), 7.

Peng, W., Shuainan, Z., Yanru, P., Shuchao, C., & Yuhu, Z. (2021). Estimation of photovoltaic power generation potential in 2020 and 2030 using land resource changes: An empirical study from China. *Energy, 219*. 119611

Pengyu, L., Jie, P., Han, G., Haifeng, T., Huajun, F., & Li, W. (2022). Evaluating the Accuracy and Spatial Agreement of Five Global Land Cover Datasets in the Ecologically Vulnerable South China Karst. *Remote Sensing, 14*(13), 3090.

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., . . . Rostan, F. (2012). GMES Sentinel-1 mission. *Remote Sensing of Environment, 120*, 9-24. https://doi.org/10.1016/j.rse.2011.05.028

Wei, W., Junxing, C., Guangzeng, T., Liang, C., & Deyong, K. (2021). Research on Accurate Extraction of Photovoltaic Power Station from Multi-source Remote Sensing. *Beijing Surveying and Mapping, 35*(12), 1534-1540. https://doi.org/10.19580/j.cnki.1007-3000.2021.12.008

E-ISSN: 2948-5231

Wu, A. N., & Biljecki, F. (2021). Roofpedia: Automatic mapping of green and solar roofs for an open roofscape registry and evaluation of urban sustainability. *Landscape and Urban Planning, 214*. https://doi.org/10.1016/j.landurbplan.2021.104167

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 System Science Data, 14*(8), 3743-3755. https://doi.org/10.5194/essd-14-3743-2022.

E-ISSN: 2948-5231