Decadal surge of water-surface solar in China's Yangtze Delta: A high-fidelity SAR-optical fusion inventory (2015–2024)

Yue Yan[1], Xin Jiang[1,*], Sihuan Wei[2,3,4], Yubin Jin[2,5], Xinyu Zou[1], Junwei Liu[4], Yaotong Cai[1], Jianhuai Ye[1], Zhilin Guo[1], Zhenzhong Zeng[1,*]

[1] School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China

[2] Zhejiang Key Laboratory of Industrial Intelligence and Digital Twin, Eastern Institute of Technology, Ningbo, Zhejiang, P.R. China

[3] Department of Building Environment and Energy Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China

[4] Internatiaonl Centre of Urban Energy Nexus, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China

[5] Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China

[*] *Correspondence to*: Zhenzhong Zeng (zengzz@sustech.edu) and Xin Jiang (jiangx@sustech.edu.cn)

Manuscript for *Earth System Science Data*

April 8, 2026

Abstract: China hosts approximately 97% of the world's water-surface photovoltaics (WPV), with nearly two-thirds of its national capacity concentrated in the Yangtze River Delta (YRD), a densely populated economic powerhouse facing intense land-energy trade-offs. Despite this dominance, no high-resolution, decade-long inventory has existed to track this rapid expansion. WPV detection using optical RS imagery is severely limited by persistent cloud cover, water surface reflections, and spectral confusion, compromising long-term consistency over aquatic environments. Here, we developed a multi-sensor fusion framework integrating all-weather Sentinel-1 Synthetic Aperture Radar (SAR) and annual composite Sentinel-2 optical imagery. Key features include six Sentinel-2 bands, spectral indices (NDVI, MNDWI, NDBI, NDPI, and SAVI), texture metrics, and dual-polarization SAR backscatter. We trained a Random Forest classifier on 55,849 verified samples to generate annual WPV maps for 2015-2024. Afterwards, we applied post-processing procedures, including noise removal, patch merging, and area thresholding, and further validated installation years and removed misclassified areas through manual inspection of Google Earth time-series imagery. The well-constructed dataset of the first 10 m-resolution WPV atlas for the YRD maps 401 validated projects with a cumulative area of 145.4 km^2 by 2024. It outperforms existing global PV inventories with an overall accuracy of 97.3% and a Kappa coefficient of 0.94. The results reveal rapid expansion from 17.4 km^2 in 2015 to 145.4 km^2 in 2024, with 87% deployed on natural lakes, with a marked shift in leadership from Jiangsu to Anhui, and clear spatial clustering near grid infrastructure and stable water bodies. This high-fidelity inventory provides a robust foundation for monitoring WPV evolution, assessing environmental impacts, and informing sustainable energy planning in the world's leading floating solar region.

Keywords: Water-surface photovoltaics, Yangtze River Delta, Sentinel-1/2 fusion, Time-series mapping, Random Forest classification

1 Introduction

The global shift from carbon-intensive energy systems to carbon-extensive renewables is accelerating in response to growing electricity demand and the urgent need to mitigate climate change. Solar photovoltaics (PV) dominates its growth, owing to its continued cost declines, and accounted for nearly 80% of new capacity additions in 2024 (IEA, 2025; Bogdanov *et al.*, 2021). However, the substantial land footprint required by conventional PV systems increasingly conflicts with agriculture and other critical land uses (Capellán-Pérez *et al.*, 2017; van de Ven *et al.*, 2021; Wei *et al.*, 2025). Water-surface photovoltaics (WPV) emerges as a promising alternative. Deployed on natural lakes, reservoirs, and other water bodies, WPV alleviates land-use competition while offering additional benefits, such as reduced water evaporation loss (Forester *et al.*, 2025; Pouran *et al.*, 2022; Jin *et al.*, 2023). China hosts approximately 97% of global WPV, with nearly two-thirds of its national capacity concentrated in the water-abundant, energy-intensive Yangtze River Delta (YRD) region (Chen *et al.*, 2024). For the YRD and similar regions, understanding the spatial dynamics and scalability of WPV is crucial for optimizing renewable energy strategies that balance energy security, land stewardship, and ecosystem sustainability.

In recent years, remote sensing combined with machine learning has become prevalent for extracting photovoltaic installations at regional or global scales (Zhang *et al.*, 2023; Ortiz *et al.*, 2022; Zhang *et al.*, 2022). Among these, classification algorithms such as Random Forest have demonstrated strong performance and robustness (Belgiu and Drăguţ, 2016). These techniques typically leveraged spectral, geometric, and textural features from satellite imagery to enable efficient, large-scale PV detection (Zhang *et al.*, 2021; Chen *et al.*, 2022). Building on these advances, several global photovoltaic inventories have been developed to provide large-scale datasets of PV installations. For example, A Global Inventory of Photovoltaic Installations maps commercial-, industrial-, and utility-scale PV facilities worldwide using multi-source satellite imagery, including Sentinel-2 (10 m) and high-resolution SPOT-6/7 imagery, and was produced using deep learning with manual verification for installations identified primarily from 2016 to 2018 (Kruitwagen *et al.*, 2021). Similarly, Global Renewables Watch integrates quarterly

PlanetScope imagery at 4.7 m spatial resolution with deep learning–based semantic segmentation to map global solar and wind infrastructure from 2017 Q4 to 2024 Q2, enabling regular updates and temporal tracking of infrastructure development (Robinson *et al.*, 2025). Compared with these products, our study is specifically designed for WPV mapping and combines SAR and optical observations to better distinguish floating PV from surrounding water backgrounds, while also providing annual, temporally consistent WPV inventories for 2015–2024.

Despite these advances, several challenges persist, particularly for complex aquatic environments and long-term monitoring. Optical imagery, which forms the foundation of most existing methods, is highly sensitive to weather conditions such as clouds, fog, and overcast skies, and cannot acquire data at night. Over water bodies, its performance is further constrained by reflections and shadows. Furthermore, although per-image accuracies can exceed 96% (Kruitwagen *et al.*, 2021; Xia *et al.*, 2022, 2023), small errors accumulate substantially in long-term time-series mapping. For instance, mapping with 96% annual accuracy over a decade yields only about 67% consistency across the full period ($0.96^{10} \approx 0.67$). This compounded uncertainty, effectively rendering one-third of time-series outcomes unreliable, complicates spatiotemporal assessments and often results in confusion between photovoltaic installations and water surfaces. As a result, ensuring both temporal consistency and high fidelity in long-term water-based PV mapping remains a key methodological challenge.

To directly address these limitations, this study proposes a robust and multidimensional feature-fusion framework for WPV extraction in the YRD region. Our approach utilizes annual composite Sentinel-2 imagery (2015-2024) and incorporates Sentinel-1 SAR observations to provide all-weather detection and capture strong backscatter signatures of WPV's metallic structures. This combination substantially enhances single-period mapping accuracy and reduces misclassifications. To ensure comprehensive spatial coverage, we first identified candidate WPV zones through multi-year temporal compositing and water-mask filtering, thereby constraining detection to likely water-based locations and suppressing non-water interference. Subsequently, all detected WPV regions underwent rigorous manual inspection and refinement using high-resolution Google Earth imagery. This step enabled the precise removal of false positives and the reliable determination of installation years. The resulting decade-long WPV dataset achieves high accuracy and temporal consistency, representing the most reliable long-term inventory of WPV development to date and providing a robust foundation for understanding spatiotemporal growth dynamics and policy-relevant deployment patterns.

2 Datasets and Methods

2.1 Study Area

China leads the world in WPV system development, holding approximately 96.82% of the global WPV-installed surface area in 2019, as shown in Fig. 1a. Within China, the YRD region represents a significant WPV hub, accounting for nearly two-thirds of the national total. Specifically, Jiangsu (30.21%), Anhui (19.75%), and Zhejiang (9.97%) provinces exhibit the highest installed WPV capacities (Xia *et al.*, 2022, 2023).

Located in eastern China along the lower reaches of the Yangtze River (Fig. 1b), the YRD is characterized by abundant water resources, including a dense network of rivers, lakes, and reservoirs (Fig. 1c). These natural conditions are highly favorable for large-scale WPV deployment. Moreover, as one of China's most economically developed and industrially concentrated areas, the YRD faces substantial electricity demand (Xu *et al.*, 2023). The rapid growth of WPV in this region not only helps alleviate regional energy pressures but also plays a vital role in energy restructuring and the transition toward low-carbon development. Considering its current deployment scale, water resource conditions, and future development potential, this study focuses on the YRD as its primary research area.

2.2 Datasets

2.2.1 Satellite Datasets

This study primarily utilizes Sentinel-1 Synthetic Aperture Radar (SAR) and Sentinel-2 Multispectral Instrument (MSI) imagery, both of which are accessible via the Google Earth

Engine (GEE) platform. The selected temporal coverage, from 2015 to 2024, aligns with the rapid development timeline of WPV installations in China.

Sentinel-1, equipped with a C-band SAR sensor, provides all-weather, day-and-night imaging capabilities, making it ideal for long-term dynamic monitoring as it's unaffected by cloud cover, precipitation, or illumination. We utilized the Sentinel-1 Ground Range Detected product from GEE, which features a 10-meter spatial resolution and a 6–12-day revisit time. To comprehensively capture radar backscatter characteristics of water bodies and WPV structures, Vertical Transmit/Vertical Receive (VV) and Vertical Transmit/Horizontal Receive (VH) polarization channels were selected.

Sentinel-2 provides high-resolution, multispectral optical and near-infrared imagery, well-suited for identifying and classifying WPV regions. We utilized atmospherically corrected Level-2A surface reflectance products from the Sentinel-2 MSI sensor, which provides 13 spectral bands. Six key bands sensitive to water bodies and artificial structures were selected: Bands 2–4 (visible), Band 8 (NIR), and Bands 11–12 (SWIR), all at 10-meter resolution with a 5-day revisit. To reduce cloud interference, a cloud-masking algorithm was applied, and annual median composites were generated from all available images (Gorelick *et al.*, 2017). These composites ensure radiometric consistency and provide a stable spatial baseline for dynamic WPV detection and temporal analysis.

2.2.2 Water Body Datasets

A water mask of the Yangtze River Delta (YRD) was constructed to geographically delineate the study area. To construct the water mask for WPV mapping, we integrated four waterbody-related datasets: HydroLAKES, GRanD, GOODD, and GeoDar (Lehner *et al.*, 2011; Messager *et al.*, 2016; Mulligan *et al.*, 2020; Wang *et al.*, 2022). HydroLAKES provided lake boundaries and associated attributes, while GRanD supplied reservoir polygons and reservoir-related information. GOODD and GeoDar, which contain georeferenced dam records, were used as complementary datasets to support reservoir identification where reservoir boundary or attribute information was incomplete or uncertain. These datasets were harmonized into a unified maximum historical water-extent layer for the study area. This integrated layer was used as a stable spatial mask to constrain the analysis to long-term potential water bodies, rather than as a year-specific annual water mask.

In this study, reservoirs were defined as water bodies formed by dams or other engineered hydraulic infrastructure and subject to artificial water-level regulation. In contrast, lakes were defined as inland water bodies without reservoir functions, including natural lakes as well as artificial water bodies such as aquaculture ponds and irrigation ponds. This distinction is relevant for WPV analysis because reservoirs generally exhibit more stable water levels and are associated with different ownership and management regimes than lakes. By integrating these four datasets, a foundational water layer for the study area was generated. This layer provides a stable spatial reference and serves as a validation framework for identifying water-surface photovoltaics using Sentinel-1 SAR and Sentinel-2 optical imagery.

2.2.3 Training and Validation Samples

The training and validation samples used in this study were derived from the WPV inventory (Xia *et al.*, 2022), which mapped the distribution of WPV installations in China for 2021. Because the original dataset contained several misclassified regions, a manual verification process was conducted to ensure the reliability of the labels. Each WPV polygon was visually checked using high-resolution satellite imagery, and misidentified areas were removed. WPV sample points were generated proportionally to the area of each PV polygon to ensure adequate representation of large installations. Examples from three representative WPV regions are shown in Figs. 2a–c, demonstrating typical distribution patterns across different water bodies. Non-WPV samples were randomly selected from water surfaces without PV coverage to provide balanced class representation.

The spatial distribution of all sample points is illustrated in Fig. 2d, where red points represent WPV samples and blue points represent non-PV samples. In total, the dataset consists of 55,849 labeled points (80% for training and 20% for validation), including 28,332 WPV and 27,517 non-WPV samples (Fig. 2e). These well-validated samples formed the foundation for subsequent annual classifications from 2015 to 2024.

2.3 WPV Extraction Workflow

2.3.1 WPV Feature Engineering

To reduce interference from temporary shoreline fluctuations and to focus the analysis on long-term potential aquatic areas, we filtered Sentinel-2 MSI imagery using a unified maximum historical water-extent mask, rather than deriving an independent water mask for each year. Six spectral bands (Bands 2–4 [visible], Band 8 [NIR], and Bands 11–12 [SWIR]) were selected as key features due to their sensitivity to water bodies and artificial structures. To minimize the influence of cloud cover and provide stable annual surface conditions, annual median composites were generated from all available images within each year. To enhance the distinction between water bodies and WPV installations, we integrated a comprehensive set of additional features:

Spectral indices: Normalized Difference Vegetation Index (NDVI), Modified Normalized Difference Water Index (MNDWI), Normalized Difference Built-up Index (NDBI), Normalized Difference Photovoltaic Index (NDPI), and Soil-Adjusted Vegetation Index (SAVI), based on Eqs. (1)– (5). These indices have proven effective in photovoltaic detection (Feng *et al.*, 2024). Specifically, NDVI and SAVI helped reduce confusion between WPV and vegetated surfaces such as emergent or floating vegetation near shorelines, MNDWI improved the separation of panel-covered water from open water, and NDBI/NDPI enhanced the identification of artificial panel surfaces that might otherwise resemble dark water in optical imagery.

Although some spectral indices are derived from the original optical bands and are therefore correlated with them, Random Forest is generally robust to multicollinearity among predictors in classification tasks. As a result, the inclusion of both original bands and derived indices is not expected to adversely affect model performance, although it may influence the interpretation of variable importance.

$$NDVI = \frac{\rho(NIR) - \rho(RED)}{\rho(NIR) + \rho(RED)} \tag{1}$$

$$MNDWI = \frac{\rho(GREEN) - \rho(SWIR1)}{\rho(GREEN) + \rho(SWIR1)} \tag{2}$$

$$NDBI = \frac{\rho(SWIR1) - \rho(NIR)}{\rho(SWIR1) + \rho(NIR)} \quad (3)$$

$$NDPI = \frac{\rho(SWIR1) - \rho(NIR)}{\rho(NIR) - \rho(SWIR2)} \quad (4)$$

$$SAVI = 1.5\frac{\rho(NIR) - \rho(RED)}{\rho(NIR) + \rho(RED) + 0.5} \quad (5)$$

Texture features: Calculated from the B8 (NIR) band using the Gray Level Co-occurrence Matrix (Haralick et al., 1973). These features capture the characteristic spatial patterns of WPV arrays, which typically exhibit clear, regular boundaries that contrast with natural water bodies.

SAR-based backscatter data: We incorporated annual mean values from Sentinel-1 SAR VV and VH polarization bands to provide complementary all-weather backscatter information, crucial for robust WPV identification, especially over water.

2.3.2 Annual WPV Classification

Classification was conducted using the Random Forest algorithm implemented within the Google Earth Engine (GEE) platform, selected for its robustness, computational efficiency, and demonstrated success in PV mapping tasks (Feng *et al.*, 2024; Zhang *et al.*, 2023). The Random Forest classifier was trained on a labeled sample dataset derived from both WPV and non-WPV-covered water bodies (Fig. 3a). The Random Forest classifier was implemented in Google Earth Engine using the smileRandomForest algorithm. We used 80 trees and selected 7 variables randomly at each split. The remaining hyperparameters were kept at the default settings of the GEE implementation, including minimum leaf population = 1, bag fraction = 0.5, maximum nodes = null, and seed = 0. After model training, the classifier was applied to annual imagery to perform classification for each year from 2015 to 2024, producing a decade-long time series of WPV distribution maps.

2.3.3 Automated Post-Processing

Following the initial classification, we applied a systematic post-processing methodology to refine the results, remove non-WPV areas, and consolidate adjacent patches. This process, consistent with previous work (Hirayama *et al.*, 2019), aimed to improve overall map quality and reduce the workload for subsequent visual interpretation (Fig. 3b). Given that WPV installations typically occupy relatively large areas, we first performed noise removal by identifying and eliminating classified patches with fewer than 10 pixels. Additionally, WPV arrays located in close proximity within the same water body are often part of the same project and installed concurrently. To accurately represent this, adjacent patches within a 100-meter buffer (i.e., less than 200 meters apart) were merged into single units. Finally, recognizing that larger WPV patches generally correlate with higher classification accuracy, only those with an area greater than 0.001 km^2 were retained in the final results, ensuring the inclusion of reliable WPV installations.

2.4 Accuracy Assessment and Manual Refinement

2.4.1 Classification Model Assessment

To support the classification workflow, a stratified random sample comprising 20% of the total dataset (5,667 WPV points and 5,507 non-WPV) was held out for independent accuracy assessment. Classification performance was evaluated using four standard metrics: User Accuracy (UA), Producer Accuracy (PA), Overall Accuracy (OA), and the Kappa coefficient.

2.4.2 Manual Refinement and Final Dataset Creation

To achieve the highest accuracy and completeness in our WPV extraction, we integrated the annual classified WPV maps (2015–2024) with external WPV datasets (Xia *et al.*, 2022), creating a comprehensive set of potential WPV regions. Each potential region was then interpreted and corrected using high-resolution satellite imagery from Google Earth (Fig. 3c) to accurately identify and remove misclassified non-WPV areas, thereby substantially improving the reliability of the final dataset. Specifically, each potential WPV region was checked for (1) its location within the water body, (2) the presence of regular and repetitive photovoltaic array patterns, and (3) separation from non-WPV objects such as shoreline buildings, roads, embankments, or floating vegetation. Since WPV installations are typically long-lasting, their installation year was determined by identifying the first year each site visibly appeared in high-resolution Google Earth imagery sequences. To maintain temporal consistency across the annual series, previously confirmed WPV areas were retained in subsequent years, and only newly detected regions were added. Through this temporal consistency rule, the annual WPV series followed a cumulative, non-decreasing pattern, effectively reducing spurious year-to-year disappearance caused by classification noise, short-term hydrological variations, or image-quality differences, thereby enhancing the reliability of decadal trend estimation. Finally, internal gaps within the identified WPV patches were filled to ensure spatial completeness, facilitating more precise calculations of area and surface coverage.

3 Result

3.1 Accuracy Validation and Comparison of WPV Extraction Results

To comprehensively evaluate the reliability and robustness of our WPV extraction results, we conducted a multi-level validation and comparison process. This section presents a systematic assessment of the classification performance from three complementary perspectives. First, the initial classification outputs were visually inspected and refined to evaluate the effectiveness of the post-processing procedure. Then, single-year and multi-year merged datasets were compared to examine the temporal stability of WPV detection results. Finally, our dataset was compared with existing global PV products, including a global inventory of photovoltaic solar energy generating units as of the end of 2018 (Kruitwagen et al., 2021) and Global Renewables Watch in 2024 (Robinson *et al.*, 2025), to assess spatial completeness and consistency. Both qualitative and quantitative analyses were performed, including confusion-matrix-based accuracy assessment and detailed visual interpretation across representative regions. Together, these evaluations provide a thorough verification of the accuracy, continuity, and advantages of our WPV dataset.

3.1.1 Evaluation of Initial and Refined WPV Extraction Results

To visually assess the effectiveness of our post-processing workflow, we compared the initial and refined WPV classification results (Fig. 4). The initial classification results (Figs. 4b, e, h) contained a certain level of noise and misclassification, including small isolated patches and fragmented boundaries along water edges. These inaccuracies were mainly caused by spectral confusion between WPV installations and nearby structures or floating vegetation. After systematic post-processing and manual correction (Figs. 4c, f, i), the final results exhibited much cleaner boundaries and more coherent WPV patches. The red outlines clearly delineate WPV areas, demonstrating improved spatial consistency and a reduction in false positives. This refinement process substantially enhanced the reliability of the extracted WPV maps, providing a foundation for subsequent accuracy validation and spatial analysis.

3.1.2 Comparison Between Single-Year and Multi-Year Merged Results

We compare WPV extraction results obtained from single-year imagery with those derived from merged multi-year datasets (Figs. 5, 6), providing an assessment of the benefits of temporal data integration. The results indicate that single-year classifications on both lakes and reservoirs frequently suffered from incomplete coverage (Figs. 5a–b, d–e, g–h and 6a–b, d–e, g–h), resulting in fragmented or missing WPV patches due to cloud contamination or partial image coverage. By contrast, the merged 2015–2024 dataset (Figs. 5c, f, i and 6c, f, i) offered a more consistent and complete delineation of WPV boundaries, as highlighted by the orange outlines. These improvements are particularly notable in turbid or seasonally fluctuating water bodies, where single-year imagery alone may fail to capture stable WPV features. Overall, the multi-year merging strategy substantially enhanced both classification completeness and spatial continuity, providing a robust basis for temporal analyses of WPV expansion.

3.1.3 Comparison with Existing Global PV Datasets

We compared our WPV dataset with a global inventory of photovoltaic solar energy-generating units and Global Renewables Watch in terms of statistical accuracy. Six representative locations were selected to assess the performance of our dataset relative to these global datasets (Figs. 7, 8). Visual inspection indicates that the existing global datasets exhibit notable limitations in WPV identification, frequently resulting in misclassification or omission, particularly in small-scale or spatially complex inland water bodies. Fig. 7 highlights instances of incomplete or erroneous WPV identification, whereas Fig. 8 shows multiple WPV regions that were entirely missed by the global datasets but were successfully captured in our study.

Quantitative accuracy assessment (Table 1) further confirms the superior performance of our dataset. Specifically, our WPV dataset achieved an Overall Accuracy (OA) and Kappa coefficient exceeding 0.9, while the corresponding metrics for the two global datasets were generally lower (OA: 0.825–0.818, Kappa: 0.651–0.637). These results demonstrate that our dataset significantly outperforms existing global datasets in terms of both classification accuracy and spatial completeness, validating that the integration of multi-temporal imagery with regionally representative sample training can substantially enhance the accuracy and reliability of WPV extraction.

3.2 WPV Spatiotemporal Distribution and Growth Trends

WPV projects in the study area exhibit distinct spatiotemporal distribution and growth trends. Spatially, current WPV projects are primarily concentrated in Anhui and Jiangsu provinces, which provide suitable hydrological and land-use conditions for large-scale deployment (Fig. 9a). By 2024, a total of 401 WPV projects have been identified in the YRD, covering a cumulative area of 145.5 km^2. Among these, Anhui Province hosts the largest share (68.7 km^2), accounting for 47% of the total WPV area, followed by Jiangsu (64.8 km^2) and Zhejiang (12 km^2). The inset bar chart further illustrates the proportion of WPV area relative to the total water area in each province in 2024, showing that Anhui had the highest deployment intensity (1.01%), followed by Jiangsu (0.67%) and Zhejiang (0.48%). Based on the five largest WPV projects in each province (Table 2), Jiangsu generally exhibits a larger overall project scale, with most large WPV installations located on lakes.

From a temporal perspective, WPV installations have expanded markedly between 2015 and 2024, with the total area increasing by 128 km^2 (Fig. 9b). This expansion was most pronounced during the early phase (2015–2019), which contributed approximately 59.2% of the total increase, followed by a relative slowdown during 2019–2024. At the provincial level, Anhui experienced the greatest increase in WPV area over the decade, adding 66.9 km^2, followed by Jiangsu (49.1 km^2), while Zhejiang's growth remained modest, with a total addition of only 12 km^2. Notably, the spatial evolution trajectories of WPV deployment differ significantly among the three provinces. In Jiangsu, WPV development began in the northern region and gradually expanded southward and toward the coastal areas during the early years. In contrast, early WPV projects in Anhui were also concentrated in the north, but from 2022 onward, installations rapidly expanded along the Yangtze River corridor, forming a more continuous, belt-shaped distribution. Zhejiang Province, by comparison, saw only limited WPV deployment, characterized by a short burst of growth between 2017 and 2020, with most projects clustered in its northern and western regions.

In addition to the overall spatial and temporal patterns in the YRD, WPV projects of different sizes exhibit distinct trends in quantity and growth, providing additional insights into the structural characteristics of PV development in the region (Fig. 9b). Large-scale projects (> 1.0 km^2) are relatively few, experiencing rapid growth primarily between 2015 and 2018, then stabilizing. Conversely, medium-scale (0.1–1.0 km^2) and small-scale (<0.1 km^2) projects far outnumber large ones and have grown rapidly overall. In general, WPV development has evolved from localized concentrations to broader regional deployment, and from predominantly large-scale projects to a more diversified mix of small-scale and medium-scale systems.

3.3 WPV Deployment on Lakes and Reservoirs

Different types of water bodies exhibit varying levels of adaptability to WPV deployment, which directly affects the estimation of potential installed capacity and informs rational spatial planning (Bai *et al.*, 2024; Château *et al.*, 2019). Within the study area, WPV installations are predominantly located on lakes, with only a minor proportion situated on reservoirs (Fig. 10a). Lakes host the vast majority of total capacity—exceeding 126.8 km^2 by 2024, or 87.2% of the cumulative WPV area. In contrast, reservoirs account for 12.8% (18.6 km^2) of total WPV deployment. However, the distribution across waterbody types differs considerably among provinces. Jiangsu exhibits the most pronounced disparity, with approximately 98.1% of its WPV systems deployed on lakes and only 1.9% on reservoirs. Anhui and Zhejiang demonstrate higher shares of reservoir-based WPV, at 28.9% and 20.2%, respectively. The temporal evolution of WPV deployment also reveals distinct trajectories for lakes and reservoirs (Fig. 10b). Lake-based WPV installations expanded steadily between 2015 and 2024, with a total increase of 109.4 km^2. In contrast, the development of reservoir-based WPV proceeded more slowly. Its expansion was limited between 2015 and 2019 (14.9 km^2), and following the introduction of policy restrictions in 2019, reservoir-based WPV installations declined sharply, with the total area remaining below 20 km^2. These trends underscore the growing dominance of lakes as preferred sites for WPV deployment, while reservoirs have become increasingly constrained due to operational sensitivities, fluctuating water levels, and tightening environmental regulations (General Office of the Ministry of Water Resources, 2020).

The impact of WPV systems on water bodies is multifaceted, influencing various physical properties and ecological dynamics, with the extent of WPV coverage exerting differential effects based on water body characteristics (Exley *et al.*, 2021). By integrating WPV distribution data with surface water datasets, WPV coverage was calculated for 385 water bodies across the study area (Figs. 11a, b). Coverage rates varied substantially among different size classes. In smaller water bodies (0–4 km^2), WPV coverage exhibited high variability, ranging from 0% to 100%. However, as water body size increased, WPV coverage declined sharply, falling below 50% in medium-sized water bodies and below 10% in large ones (>100 km^2). Most reservoirs covered by WPV installations are relatively small (< 4 km^2), and their overall WPV coverage is slightly lower than that of lakes of comparable size.

3.4 WPV Spatial Clustering and Driving Factors

At the regional scale (Fig. 12a), WPV projects exhibit a clear pattern of localized clustering across the study area. These projects are primarily distributed along major rivers and lakes, with project sizes varying considerably, from less than 0.1 km^2 to over 3 km^2. Notably, several large, high-density clusters are observed in parts of Jiangsu and Anhui provinces (Fig. 12b-12d).

This spatial aggregation is shaped by a combination of natural and socio-economic factors. Natural conditions provide fundamental support for WPV deployment, with large water bodies having minimal fluctuations in water level and stable water quality being particularly suitable (Woolway *et al.*, 2024). For example, Sanlihe Reservoir (Fig. 12e), Gaoyou Lake (Fig. 12f), and Xizi Lake (Fig. 12g) possess stable, expansive water surfaces and favorable hydrological conditions, supporting extensive WPV arrays. Grid accessibility is another key driver influencing the spatial distribution of WPV systems (Essak and Ghosh, 2022). In both Gaoyou

Lake and Xizi Lake, WPV installations are located near urban settlements and existing power infrastructure, which facilitates grid connection and reduces energy transmission losses and associated costs. This locational advantage has made such water bodies prime targets for large-scale WPV deployment.

4 Discussion

4.1 Major Findings and Contributions

This study constructs a high-precision, decade-long (2015–2024) WPV dataset for the YRD region, thereby significantly advancing remote sensing-based WPV mapping. We specifically address key limitations in existing time-series mapping approaches, namely the constraints of optical imagery (e.g., cloud cover and water surface reflection) and the substantial error accumulation in long-term mapping (e.g., per-period accuracy of ~0.96 may degrade to ~0.67 over ten years). Our robust solution—fusing Sentinel-1 SAR data with multi-temporal Sentinel-2 imagery—effectively leverages SAR's all-weather imaging capability and the strong backscatter signals uniquely associated with the metallic structure of WPV panels, thereby significantly enhancing single-period mapping accuracy and reducing misclassification. In addition, through probabilistic filtering and meticulous manual verification using high-resolution Google Earth imagery over potential WPV regions, we achieved high accuracy and temporal consistency for this long-term dataset.

Based on this reliable dataset, we conducted the first systematic and fine-scale decadal analysis of the spatial-temporal distribution and growth trends of WPV in the YRD. We revealed that the total WPV area expanded rapidly from 17.4 km^2 in 2015 to 145.4 km^2 in 2024, along with a shift in the leading province of development from Jiangsu to Anhui. The results also detail the diverse growth trajectories of large-, medium-, and small-scale WPV projects, demonstrating an evolution from localized concentration to broader regional deployment.

4.2 Implications and Potential Applications

The high-precision, decade-long WPV dataset developed in this study has both practical and scientific relevance. By quantifying the coverage, spatial distribution, and waterbody-specific deployment characteristics of WPV installations, the dataset provides a spatial basis for WPV planning, site selection, and comparative analysis across waterbody types and coverage levels. High-coverage configurations, including those associated with "fishing-solar complementarity" systems, may indicate both high energy-generation intensity and stronger local hydrological or ecological influence (Pringle *et al.*, 2017). Such settings may be associated with greater modification of light availability, air–water exchange, evaporation, and aquatic habitat conditions. From an energy-planning perspective, they also represent an intensive deployment mode, although future expansion should account for environmental carrying capacity and site-specific management constraints (Bai *et al.*, 2024; Château *et al.*, 2019).

Beyond inventory mapping, the dataset is also relevant to broader Earth system science applications. As a spatially explicit record of water-surface solar deployment, it can support analyses of energy–water–environment interactions by linking renewable-energy expansion with waterbody functions, hydrological conditions, and ecological constraints. In this sense, the dataset may serve as a useful spatial input for regional sustainability assessments and integrated analyses of low-carbon transitions, infrastructure planning, and trade-offs among energy generation, water use, and ecosystem management.

The dataset may also support investigation of the environmental effects of WPV deployment. Its high-resolution delineation of WPV-covered and uncovered areas within the same water body enables spatially explicit comparisons between shaded and unshaded zones, as well as before-and-after analyses based on installation timing. When combined with other observations, these maps may support assessment of possible water-surface temperature differences associated with partial shading using thermal infrared products, as well as potential changes in water optical properties and algal dynamics using water-color or water-quality indicators such as chlorophyll-a, turbidity, or algal bloom proxies (Chen *et al.*, 2025; Chu and He, 2023). These applications may be particularly relevant for identifying ecological risks associated with high-coverage installations in small water bodies and for informing more ecologically responsible deployment strategies (Sahu *et al.*, 2016; Armstrong *et al.*, 2020;

Nobre *et al.*, 2023; Ma and Liu, 2022). In addition, the dataset can facilitate field validation by enabling targeted selection of representative WPV sites for in situ investigation. It should be noted, however, that the SAR–optical fusion framework is primarily designed to detect WPV extent, boundaries, and deployment timing. Although these attributes are relevant for environmental exposure assessment, the framework does not directly quantify thermal or biogeochemical responses, which require dedicated thermal infrared, water-quality remote sensing, or field observations.

4.3 Limitations and Future Research

Despite constructing a high-precision dataset, this study has several limitations. First, uncertainties in the underlying waterbody datasets remain a challenge; small water bodies (e.g., ponds) may be omitted, while imprecise boundaries and spectral confusion with nearby buildings or bare soil can cause both omissions and misclassifications at water edges (Valerio *et al.*, 2024; Wang *et al.*, 2022). Second, the diversity in WPV installation methods and structural designs introduces variations in spectral and textural characteristics, posing challenges for consistent extraction (Shi *et al.*, 2023). The current framework is expected to be reliable for stationary PV installations and for anchored, medium- to large-scale FPV systems, which represent the dominant commercial deployment type in the study region. However, uncertainty may increase for small, loosely arranged, or highly mobile floating platforms, especially where local displacement alters patch boundaries or weakens texture regularity. Finally, the area threshold applied during post-processing to eliminate noise likely leads to an underestimation of the total WPV area, particularly by excluding small-scale systems on rural ponds or aquaculture facilities (Iqra *et al.*, 2024).

In light of these limitations, future research can focus on three key directions. Methodologically, the integration of deep learning methods (e.g., U-Net, DeepLabV3+) holds promise for enhancing accuracy, particularly in delineating complex boundaries and detecting morphologically diverse or small-scale targets (Chen *et al.*, 2018; Ronneberger *et al.*, 2015). Spatially, expanding the study to national or global scales would reveal macro-level deployment trends driven by policy and market dynamics, offering valuable guidance for macro-level energy planning. Thematically, future work should place greater emphasis on evaluating the ecological impacts of WPY systems by integrating remote sensing retrievals with in-situ monitoring data, thereby promoting a coordinated approach to renewable energy development and ecosystem conservation.

5 Conclusion

This study precisely mapped the spatiotemporal distribution of Water-surface photovoltaic (WPV) systems in China's Yangtze River Delta from 2015 to 2024. Using a robust framework that integrated multi-temporal Sentinel-1 SAR and Sentinel-2 optical imagery with Random Forest classification, refined by post-processing and manual verification, we generated a high-resolution, decade-long dataset that overcomes optical imagery limitations and cumulative errors in long-term monitoring. Our findings reveal WPV's significant expansion (17.4 km^2 in 2015 to 145.4 km^2 in 2024), with deployment shifting towards Anhui. Most projects are now small- to medium-scale, primarily on lakes, exhibiting clear spatial clustering influenced by water conditions and grid access. This comprehensive WPV mapping serves as a critical data source for assessing development potential, guiding renewable energy planning, and evaluating ecological impacts. Its generalizable methodology offers a strong foundation for broader WPV monitoring and future research integrating advanced remote sensing and ecological analysis.

Acknowledgments

This study was supported by the National Natural Science Foundation of China - United Nations Environment Programme (NSFC-UNEP, no. 42361144001), National Natural Science Foundation of China (no. 42071022), Shenzhen Science and Technology Project for Sustainable Development in Special Innovation (no. KCXFZ20230731093403008), Guangdong Basic and Applied Basic Research Foundation (no. 2022A1515240070), the Shenzhen Key Laboratory of Precision Measurement and Early Warning Technology for Urban Environmental Health Risks (no. ZDSYS20220606100604008) and the start-up and high-level special funds provided by the Southern University of Science and Technology (no. 29/Y01296602; 29/Y01296122; 29/Y01296222; G030290001). This work is supported with the High Performance Computing Centers (Taiyi) at SUSTech.

Data sharing statement

All datasets used in this study are publicly accessible. The Yangtze River Delta Water-Surface Photovoltaics Dataset (2015–2024) is archived at Zenodo: https://doi.org/10.5281/zenodo.17484488. For further details or data-related inquiries, please contact the corresponding authors, Zhenzhong Zeng (zengzz@sustech.edu.cn) and Xin Jiang (jiangx@sustech.edu.cn).

CRediT authorship contribution statement

Yue Yan contributed to conceptualization, data curation, formal analysis, visualization, and writing – original draft. Zhenzhong Zeng was responsible for funding acquisition, project administration, supervision, and writing – review & editing, serving as corresponding author. Xin Jiang provided supervision, methodology guidance, and writing – review & editing. All other authors contributed to investigation, validation, and writing – review & editing.

Declaration of competing interest

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

References

Armstrong, A., Page, T., Thackeray, S. J., Hernandez, R. R., and Jones, I. D.: Integrating environmental understanding into freshwater floatovoltaic deployment using an effects hierarchy and decision trees, Environ. Res. Lett., 15, 114055, https://doi.org/10.1088/1748-9326/abbf7b, 2020.

Bai, B., Xiong, S., Ma, X., and Liao, X.: Assessment of floating solar photovoltaic potential in China, Renew. Energy, 220, 119572, https://doi.org/10.1016/j.renene.2023.119572, 2024.

Belgiu, M., and Drăguţ, L.: Random forest in remote sensing: A review of applications and future directions, ISPRS J. Photogramm. Remote Sens., 114, 24 – 31, https://doi.org/10.1016/j.isprsjprs.2016.01.011, 2016.

Bogdanov, D. et al.: Low-cost renewable electricity as the key driver of the global energy transition towards sustainability, Energy, 227, 120467, https://doi.org/10.1016/j.energy.2021.120467, 2021.

Capellán-Pérez, I., De Castro, C., and Arto, I.: Assessing vulnerabilities and limits in the transition to renewable energies: Land requirements under 100% solar energy scenarios, Renew. Sustain. Energy Rev., 77, 760–782, https://doi.org/10.1016/j.rser.2017.03.137, 2017.

Château, P.-A., Wunderlich, R. F., Wang, T.-W., Lai, H.-T., Chen, C.-C., and Chang, F.-J.: Mathematical modeling suggests high potential for the deployment of floating photovoltaic on fish ponds, Sci. Total Environ., 687, 654 – 666, https://doi.org/10.1016/j.scitotenv.2019.05.420, 2019.

Chen, D., Peng, Q., Lu, J., Huang, P., Liu, Y., and Peng, F.: Assessing effect of water photovoltaics on nearby water surface temperature using remote sensing techniques, Adv. Space Res., 75, 138–147, https://doi.org/10.1016/j.asr.2024.08.040, 2025.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H.: Encoder-Decoder with Atrous

Separable Convolution for Semantic Image Segmentation, Comput. Vis. ECCV, 2018, 833–851, https://doi.org/10.1007/978-3-030-01234-2_49, 2018.

Chen, Y., Zhou, J., Ge, Y., and Dong, J.: Uncovering the rapid expansion of photovoltaic power plants in China from 2010 to 2022 using satellite data and deep learning, Remote Sens. Environ., 305, 114100, https://doi.org/10.1016/j.rse.2024.114100, 2024.

Chen, Z., Kang, Y., Sun, Z., Wu, F., and Zhang, Q.: Extraction of photovoltaic plants using machine learning methods: A case study of the pilot energy city of Golmud, China, Remote Sens., 14, 12697, https://doi.org/10.3390/rs14112697, 2022.

Chu, H.-J., and He, Y.-C.: Remote sensing water quality inversion using sparse representation: Chlorophyll-a retrieval from Sentinel-2 MSI data, Remote Sens. Appl. Soc. Environ., 31, 101006, https://doi.org/10.1016/j.rsase.2023.101006, 2023.

Essak, L., and Ghosh, A.: Floating photovoltaics: A review, Clean Technol., 4, 466–493, https://doi.org/10.3390/cleantechnol4030046, 2022.

Exley, G., Armstrong, A., Page, T., and Jones, I. D.: Floating photovoltaics could mitigate climate change impacts on water body temperature and stratification, Sol. Energy, 219, 24–33, https://doi.org/10.1016/j.solener.2021.01.076, 2021.

Feng, Q., Niu, B., Ren, Y., Su, S., Wang, J., Shi, H., Yang, J., and Han, M.: A 10-m national-scale map of ground-mounted photovoltaic power stations in China of 2020, Sci. Data, 11, 198, https://doi.org/10.1038/s41597-024-02994-x, 2024.

Forester, E., Levin, M. O., Thorne, J. H., Armstrong, A., Pasquale, G., Vincenza Di Blasi, M. L., Scott, T. A., and Hernandez, R. R.: Siting considerations for floating solar photovoltaic energy: A systematic review, Renew. Sustain. Energy Rev., 211, 115360, https://doi.org/10.1016/j.rser.2025.115360, 2025.

General Office of the Ministry of Water Resources of the People's Republic of China: Notice on Further Strengthening the Management of Construction Projects within River and Lake Management Scope (Ban He Hu [2020] No. 177), General Office of the Ministry of Water Resources, Beijing, http://www.mwr.gov.cn/zwgk/gknr/202510/t20251001_2079778.html (last access: 1

April 2026), 2020.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R.: Google Earth Engine: Planetary-scale geospatial analysis for everyone, Remote Sens. Environ., 202, 18–27, https://doi.org/10.1016/j.rse.2017.06.031, 2017.

Haralick, R. M., Shanmugam, K., and Dinstein, I.: Textural features for image classification, IEEE Trans. Syst. Man Cybern., SMC-3, 610 – 621, https://doi.org/10.1109/TSMC.1973.4309314, 1973.

Hirayama, H., Sharma, R. C., Tomita, M., and Hara, K.: Evaluating multiple classifier system for the reduction of salt-and-pepper noise in the classification of very-high-resolution satellite images, Int. J. Remote Sens., 40, 2542 – 2557, https://doi.org/10.1080/01431161.2018.1528400, 2019.

International Energy Agency (IEA): Global Energy Review 2025, IEA, Paris, https://www.iea.org/reports/global-energy-review-2025, 2025, Licence: CC BY 4.0.

Iqra, G., Giri, K. J., and Javed, M.: Small object detection in diverse application landscapes: A survey, Multimed. Tools Appl., 83, 88645–88680, https://doi.org/10.1007/s11042-024-18866-w, 2024.

Jin, Y., Hu, S., Ziegler, A. D., Gibson, L., Campbell, J. E., Xu, R., Chen, D., Zhu, K., Zheng, Y., Ye, B., Ye, F., and Zeng, Z: Energy production and water savings from floating solar photovoltaics on global reservoirs, Nat. Sustain., 6, 865 – 874, https://doi.org/10.1038/s41893-023-01089-6, 2023.

Kruitwagen, L., Story, K. T., Friedrich, J., Byers, L., Skillman, S., and Hepburn, C.: A global inventory of photovoltaic solar energy generating units, Nature, 598, 604 – 610, https://doi.org/10.1038/s41586-021-03957-7, 2021.

Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J. C., Rödel, R., Sindorf, N., and Wisser, D.: High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management, Front. Ecol. Environ., 9, 494 – 502, https://doi.org/10.1890/100125, 2011.

Ma, C., and Liu, Z.: Water-surface photovoltaics: Performance, utilization, and interactions with water eco-environment, Renew. Sustain. Energy Rev., 167, 112823, https://doi.org/10.1016/j.rser.2022.112823, 2022.

Messager, M. L., Lehner, B., Grill, G., Nedeva, I., and Schmitt, O.: Estimating the volume and age of water stored in global lakes using a geo-statistical approach, Nat. Commun., 7, 1–11, https://doi.org/10.1038/ncomms13603, 2016.

Mulligan, M., van Soesbergen, A., and Sáenz, L.: GOODD, a global dataset of more than 38,000 georeferenced dams, Sci. Data, 7, 1–8, https://doi.org/10.1038/s41597-020-0362-5, 2020.

Nobre, R., Boulêtreau, S., Colas, F., Azemar, F., Tudesque, L., Parthuisot, N., Favriou, P., and Cucherousset, J.: Potential ecological impacts of floating photovoltaics on lake biodiversity and ecosystem functioning, Renew. Sustain. Energy Rev., 188, 113852, https://doi.org/10.1016/j.rser.2023.113852, 2023.

Ortiz, A., Negandhi, D., Mysorekar, S. R., Nagaraju, S. K., Kiesecker, J., Robinson, C., Bhatia, P., Khurana, A., Wang, J., Oviedo, F., and Ferres, J. L.: An Artificial Intelligence Dataset for Solar Energy Locations in India, Sci. Data, 9, 497, https://doi.org/10.1038/s41597-022-01499-9, 2022.

Pouran, H. M., Padilha Campos Lopes, M., Nogueira, T., Alves Castelo Branco, D., and Sheng, Y.: Environmental and technical impacts of floating photovoltaic plants as an emerging clean energy technology, iScience, 25, 105253, https://doi.org/10.1016/j.isci.2022.105253, 2022.

Pringle, A. M., Handler, R. M., and Pearce, J. M.: Aquavoltaics: Synergies for dual use of water area for solar photovoltaic electricity generation and aquaculture, Renew. Sustain. Energy Rev., 80, 572–584, https://doi.org/10.1016/j.rser.2017.05.191, 2017.

Robinson, C., Ortiz, A., Kim, A., Dodhia, R., Zolli, A., Nagaraju, S. K., Oakleaf, J., Kiesecker, J., and Ferres, J. M. L.: Global Renewables Watch: A temporal dataset of solar and wind energy derived from satellite imagery, arXiv, arXiv:2503.14860, https://doi.org/10.48550/arXiv.2503.14860, 2025.

Ronneberger, O., Fischer, P., and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, arXiv preprint, arXiv:1505.04597, https://doi.org/10.48550/arXiv.1505.04597, 2015.

Sahu, A., Yadav, N., and Sudhakar, K.: Floating photovoltaic power plant: A review, Renew. Sustain. Energy Rev., 66, 815–824, https://doi.org/10.1016/j.rser.2016.08.051, 2016.

Shi, W., Yan, C., Ren, Z., Yuan, Z., Liu, Y., Zheng, S., Li, X., and Han, X.: Review on the development of marine floating photovoltaic systems, Ocean Eng., 286, 115560, https://doi.org/10.1016/j.oceaneng.2023.115560, 2023.

Valerio, F., Godinho, S., Ferraz, G., Pita, R., Gameiro, J., Silva, B., Marques, A. T., and Silva, J. P.: Multi-temporal remote sensing of inland surface waters: A fusion of Sentinel-1 & Sentinel-2 data applied to small seasonal ponds in semiarid environments, Int. J. Appl. Earth Obs. Geoinf., 135, 104283, https://doi.org/10.1016/j.jag.2024.104283, 2024.

van de Ven, D.-J., Capellán-Pérez, I., Arto, I., Cazcarro, I., De Castro, C., Patel, P., and Gonzalez-Eguino, M.: The potential land requirements and related land use change emissions of solar energy, Sci. Rep., 11, 2907, https://doi.org/10.1038/s41598-021-82042-5, 2021.

Wang, J., Walter, B. A., Yao, F., Song, C., Ding, M., Maroof, A. S., Zhu, J., Fan, C., McAlister, J. M., Sikder, S., Sheng, Y., Allen, G. H., Crétaux, J.-F., and Wada, Y.: GeoDAR: Georeferenced global dams and reservoirs dataset for bridging attributes and geolocations, Earth Syst. Sci. Data, 14, 1869–1899, https://doi.org/10.5194/essd-14-1869-2022, 2022.

Wang, L., Bie, W., Li, H., Liao, T., Ding, X., Wu, G., and Fei, T.: Small water body detection and water quality variations with changing human activity intensity in Wuhan, Remote Sens., 14, 200, https://doi.org/10.3390/rs14010200, 2022.

Wei, S., Chen, Y., and Zeng, Z.: An unexpectedly large proportion of photovoltaic facilities installed on cropland, Innov. Energy, 2, 100070 – 100073, https://doi.org/10.59717/j.xinn-energy.2024.100070, 2025.

Woolway, R. I., Zhao, G., Rocha, S. M. G., Thackeray, S. J., and Armstrong, A.: Decarbonization potential of floating solar photovoltaics on lakes worldwide, Nat. Water, 2, 566–576, https://doi.org/10.1038/s44221-024-00251-4, 2024.

Xia, Z., Li, Y., Guo, X., and Chen, R.: High-resolution mapping of water photovoltaic development in China through satellite imagery, Int. J. Appl. Earth Obs. Geoinf., 107, 102707, https://doi.org/10.1016/j.jag.2022.102707, 2022.

Xia, Z., Li, Y., Guo, S., Chen, R., Zhang, W., Zhang, P., and Du, P.: Mapping global water-surface photovoltaics with satellite images, Renew. Sustain. Energy Rev., 187, 113760, https://doi.org/10.1016/j.rser.2023.113760, 2023.

Xu, C., Xie, D., Gu, C., Zhao, P., Wang, X., and Wang, Y.: Sustainable development pathways for energies in Yangtze River Delta urban agglomeration, Sci. Rep., 13, 18135, https://doi.org/10.1038/s41598-023-44727-x, 2023.

Zhang, H., Tian, P., Zhong, J., Liu, Y., and Li, J.: Mapping photovoltaic panels in coastal China using Sentinel-1 and Sentinel-2 images and Google Earth Engine, Remote Sens., 15, 3712, https://doi.org/10.3390/rs15153712, 2023.

Zhang, X., Xu, M., Wang, S., Huang, Y., and Xie, Z.: Mapping photovoltaic power plants in China using Landsat, random forest, and Google Earth Engine, Earth Syst. Sci. Data, 14, 3743–3755, https://doi.org/10.5194/essd-14-3743-2022, 2022.

Zhang, X., Zeraatpisheh, M., Rahman, M. M., Wang, S., and Xu, M.: Texture is important in improving the accuracy of mapping photovoltaic power plants: A case study of Ningxia Autonomous Region, China, Remote Sens., 13, 3909, https://doi.org/10.3390/rs13193909, 2021.

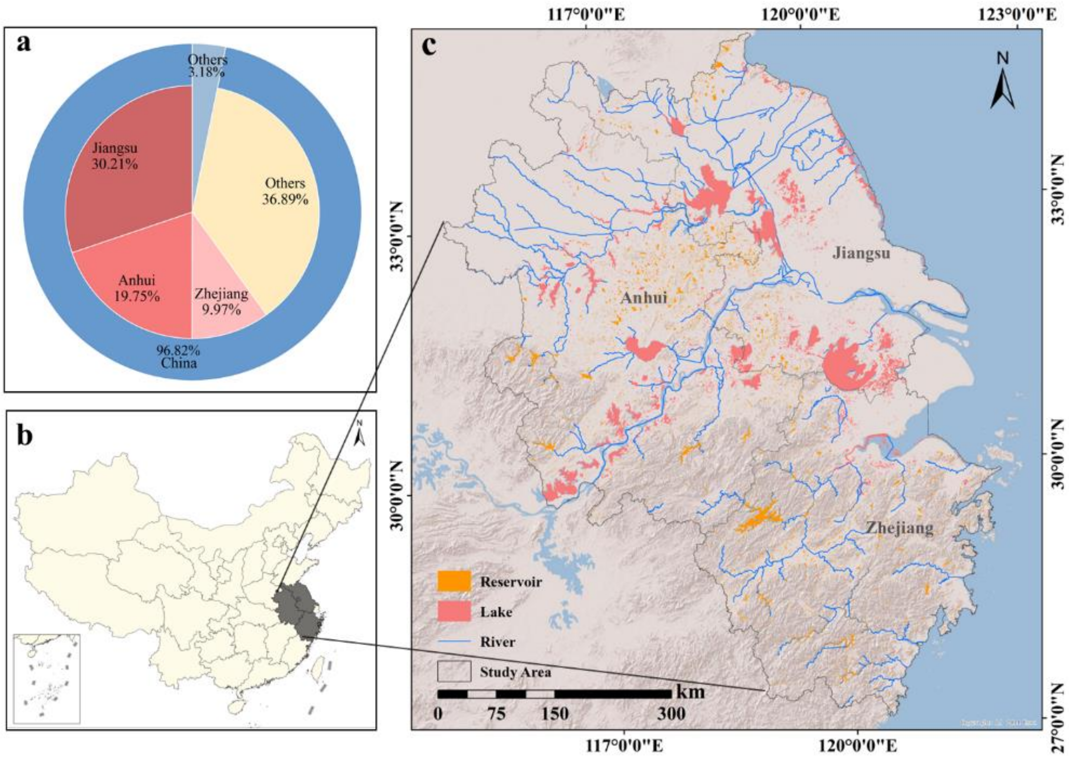
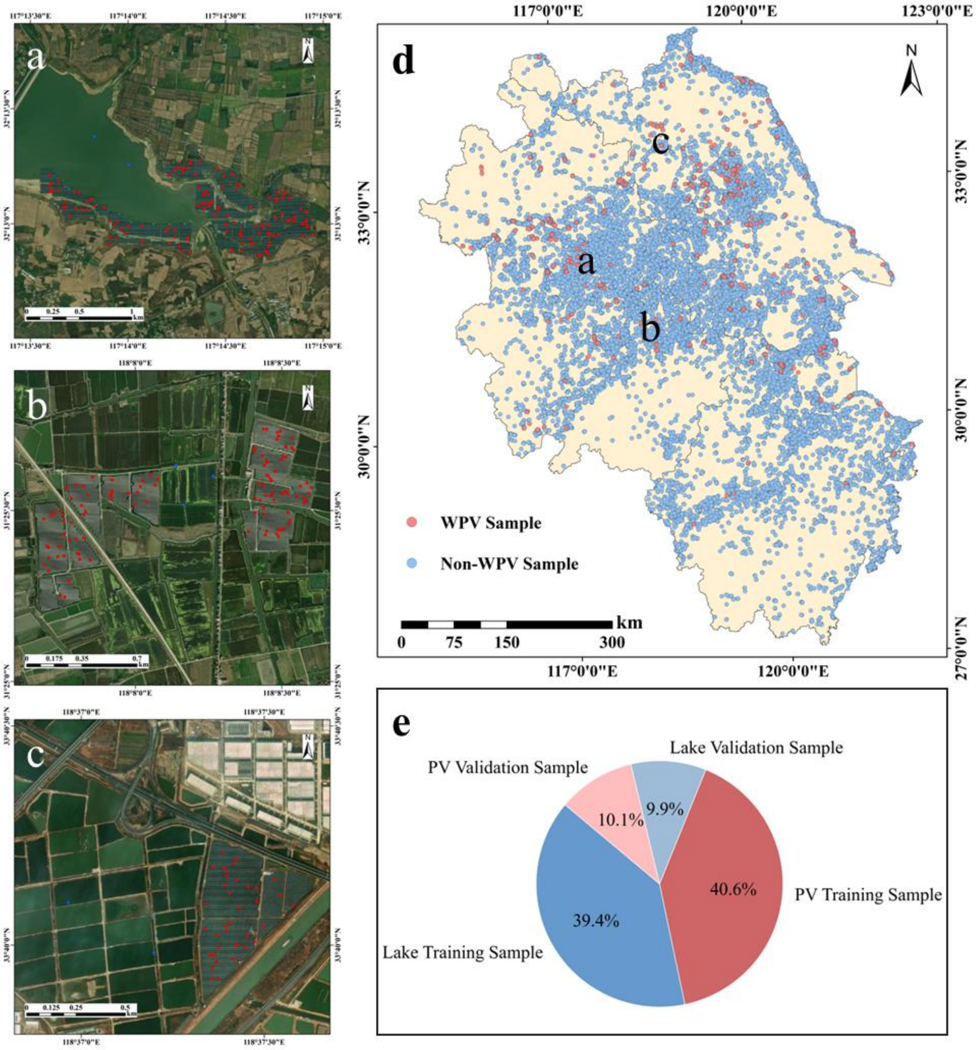


Figure 1. WPV distribution and overview of the study area. a: Proportions of WPV area in the study area relative to China, with specific contributions from Jiangsu, Anhui, and Zhejiang provinces; b: Location of the study area within China; c: Spatial distribution of reservoirs, lakes, and rivers in the study area (source: © 2014 Esri).

690



691 **Figure 2. Distribution and examples of samples.** a–c: Examples of typical sample regions,
692 where red points indicate WPV samples and blue points indicate non-WPV samples; d: Spatial
693 distribution of WPV and non-WPV samples across the study area; e: Proportions of different
694 sample categories (source: Google Earth Pro, © 2025 Airbus, © 2025 Airbus / CNES, © 2025
695 Maxar Technologies).

696

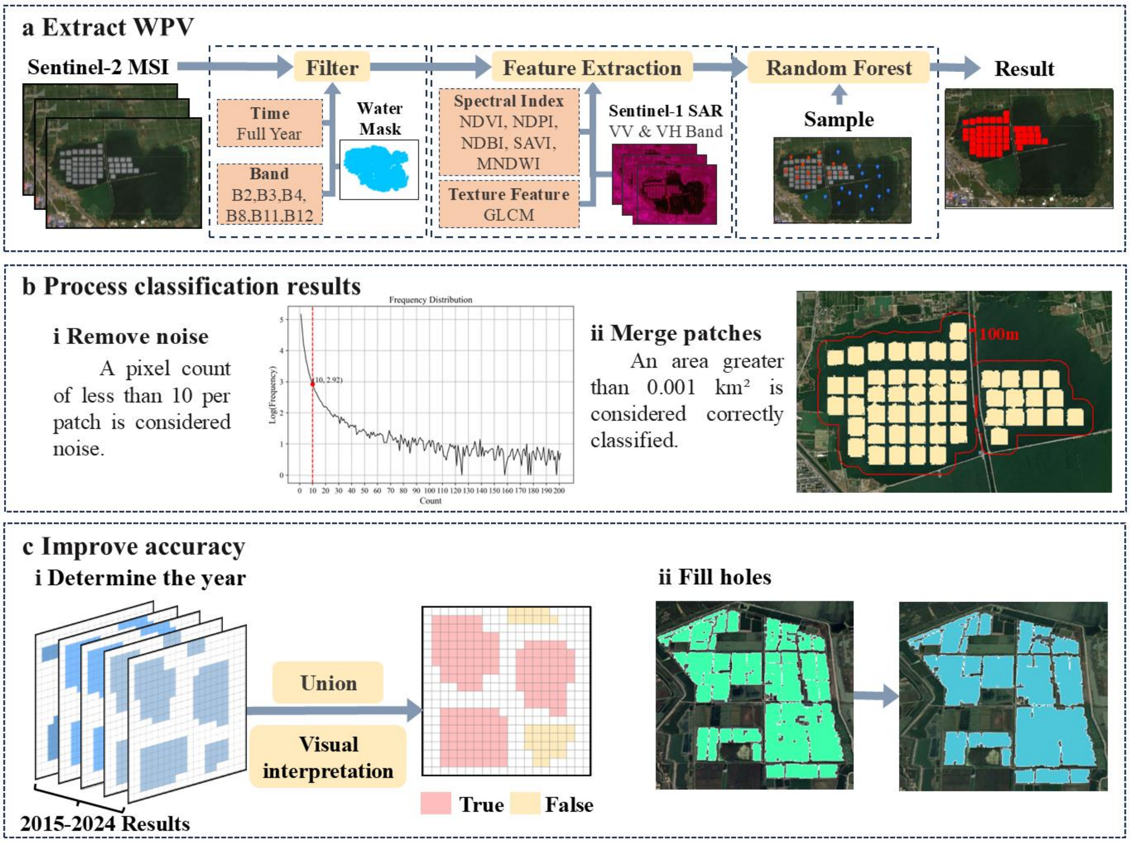


Figure 3. Flowchart of the method for extracting WPV from satellite imagery. a: WPV extraction process based on Sentinel-2 MSI and Sentinel-1 SAR data, including water masking, feature extraction (spectral indices, texture features, SAR bands), and classification using a Random Forest model; b: Post-processing of classification results: i: removing noise patches smaller than 10 pixels, and ii: merging patches within 200 meters of each other and retaining those with an area greater than 0.0001 km^2; c: Accuracy improvement strategies: i: determining the year of installation using annual union and visual interpretation, and ii: filling small holes in classified patches. The final dataset covers the period from 2015 to 2024 (source: Google Earth Pro, © 2025 Airbus, © 2025 Airbus / CNES, © 2025 Maxar Technologies).

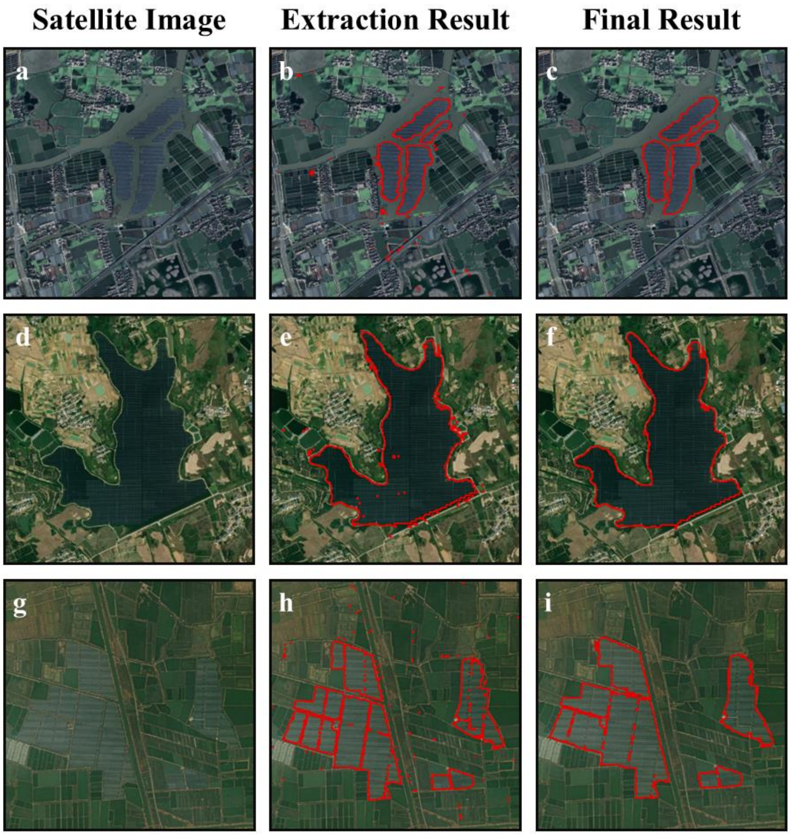


Figure 4. Examples of WPV extraction results. Panels a–c, d–f, and g–i show three representative areas. a, d, g: The satellite images; b, e, h: The initial extraction results; c, f, i: The final results after manual correction. Red outlines indicate the boundaries of WPV areas. The three areas are located at a–c: 30°51′40.54″N, 120°44′30.47″E; d–f: 32°06′03.69″N, 117°33′00.10″E; g–i: 32°58′06.10″N, 119°37′09.47″E (source: Google Earth Pro, © 2025 Airbus, © 2025 Airbus / CNES, © 2025 Maxar Technologies).

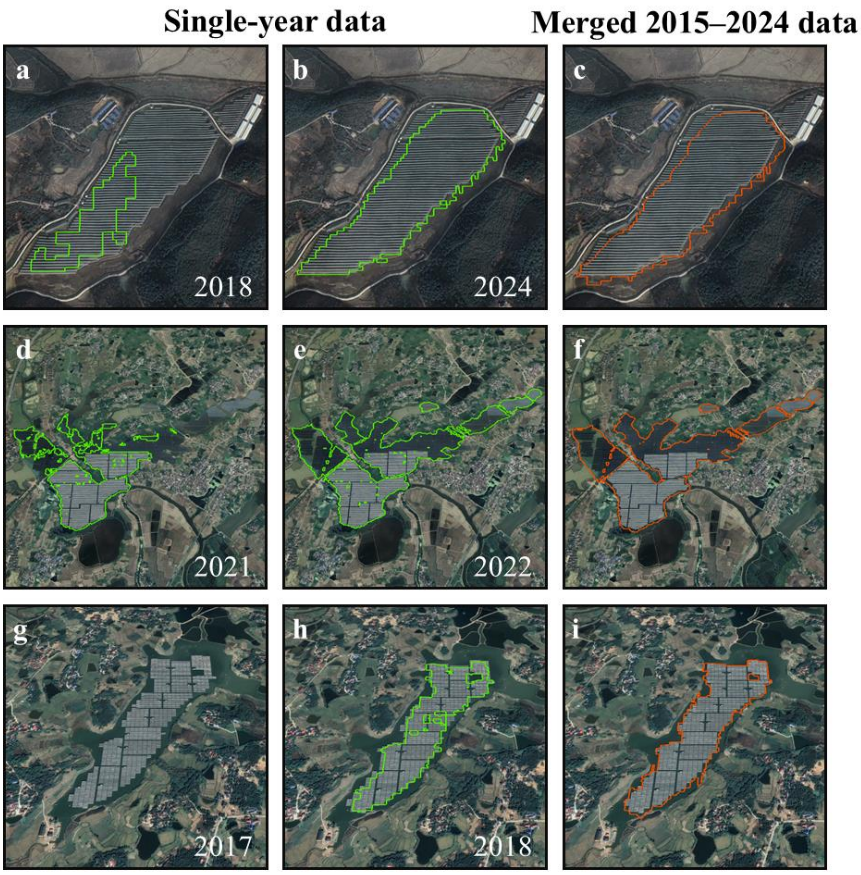


Figure 5. Comparison between single-year and merged multi-year data for WPV extraction on lakes. Panels a–c, d–f, and g–i are three representative areas. a, d, g, and b, e, h: WPV extraction results from single-year images of different years, outlined in green; c, f, i: The extraction results based on the merged 2015–2024 dataset, outlined in orange. The three areas are located at a–c: 30°22′19.59″N, 116°21′50.00″E; d–f: 30°36′26.69″N, 117°17′45.92″E; and g–i: 30°44′55.34″N, 116°57′05.93″E (source: Google Earth Pro, © 2025 Airbus, © 2025 Airbus / CNES, © 2025 Maxar Technologies).

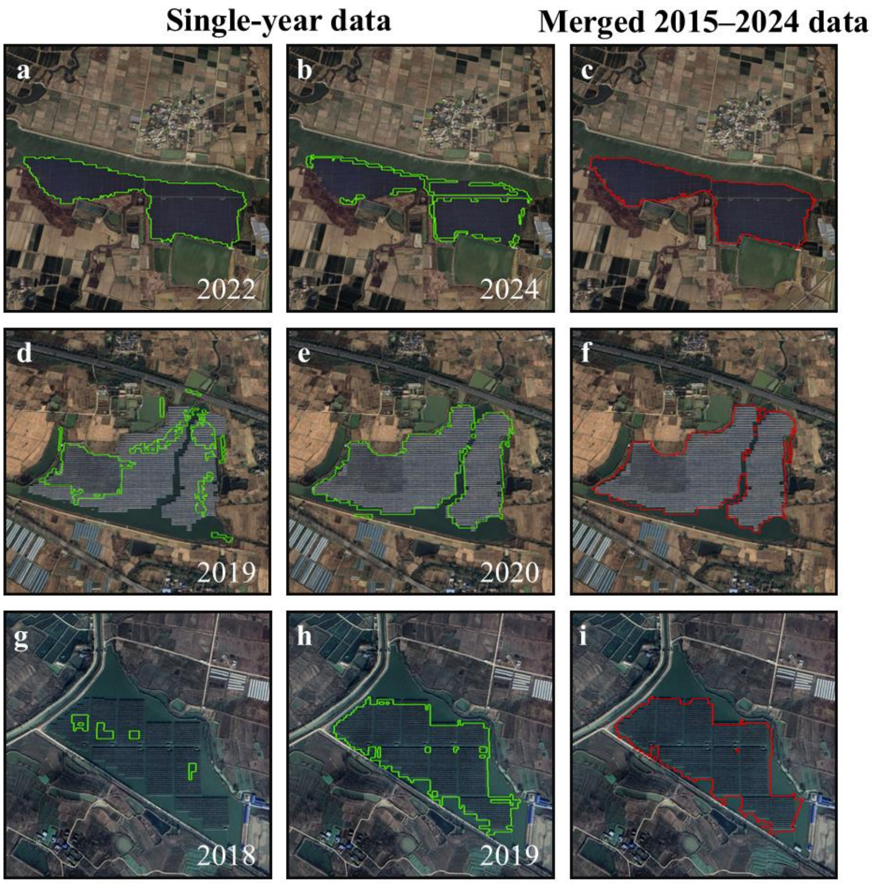


Figure 6. Comparison between single-year and merged multi-year data for WPV extraction on reservoirs. Panels a–c, d–f, and g–i are three representative areas. a, d, g, and b, e, h: WPV extraction results from single-year images of different years, outlined in green; c, f, i: The extraction results based on the merged 2015–2024 dataset, outlined in red. The three areas are located at a–c: 32°43′54.10″N, 117°41′35.43″E; d–f: 32°33′04.04″N, 116°55′38.40″E; and g–i: 32°11′46.56″N, 117°02′51.31″E (source: Google Earth Pro, © 2025 Airbus, © 2025 Airbus / CNES, © 2025 Maxar Technologies).

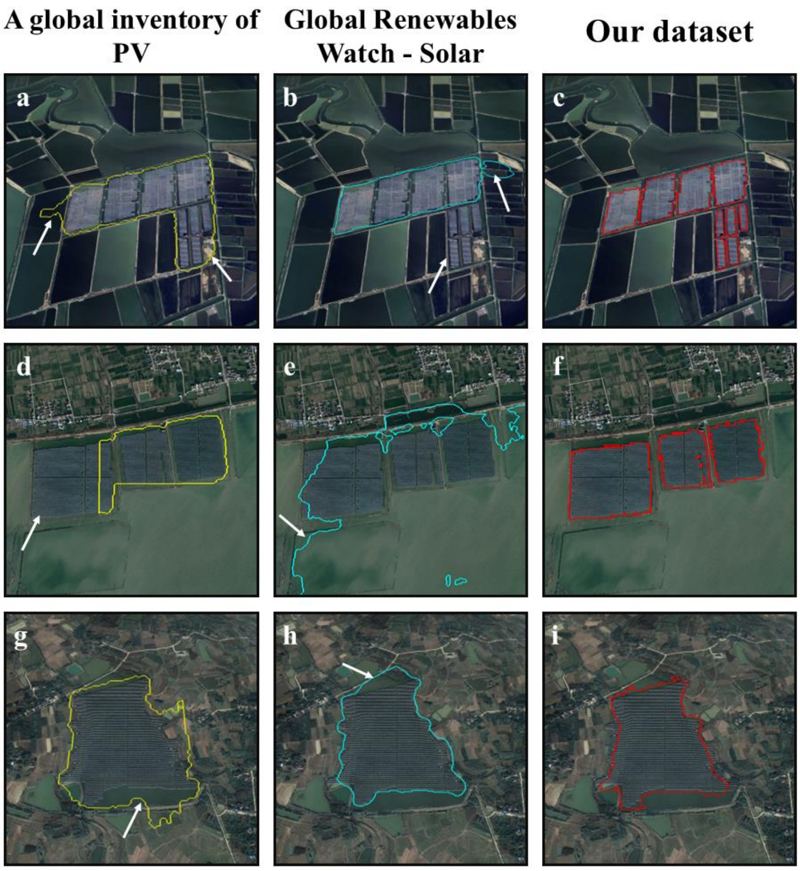


Figure 7. Comparison of our WPV extraction results with two global PV datasets (Kruitwagen *et al.*, 2021; Robinson *et al.*, 2025): examples of incomplete and incorrect identification. Panels a–c, d–f, and g–i are three representative areas. a, d, g: A global inventory of PV Kruitwagen *et al.*, 2021, outlined in yellow; b, e, h: Global Renewables Watch, 2023, outlined in blue; c, f, i: Our extraction results, outlined in red. The three areas are located at a–c: 31°07′51.26″N, 119°02′42.41″E; d–f: 32°36′34.53″N, 116°34′24.75″E; g–i: 31°35′19.28″N, 117°06′36.42″E (source: Google Earth Pro, © 2025 Airbus, © 2025 Airbus / CNES, © 2025 Maxar Technologies).

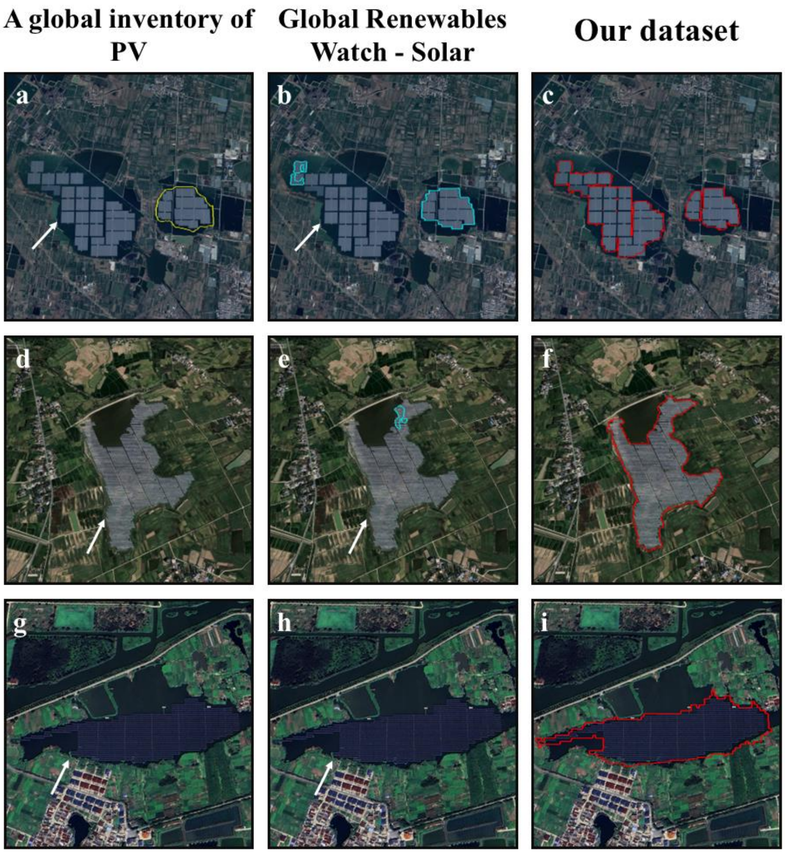


Figure 8. Comparison of our WPV extraction results with two global PV datasets (Kruitwagen *et al.*, 2021; Robinson *et al.*, 2025): examples of undetected FPV installations. Panels a–c, d–f, and g–i are three representative areas. a, d, g: A global inventory of PV Kruitwagen *et al.*, 2021, outlined in yellow; b, e, h: Global Renewables Watch, 2023, outlined in blue; c, f, i: Our extraction results, outlined in red. The three representative lake areas are located at a–c: 32°49′16.10″N, 116°49′57.69″E; d–f: 32°20′36.04″N, 117°21′11.29″E; g–i: 32°34′46.60″N, 119°58′03.91″E (source: Google Earth Pro, © 2025 Airbus, © 2025 Airbus / CNES, © 2025 Maxar Technologies).

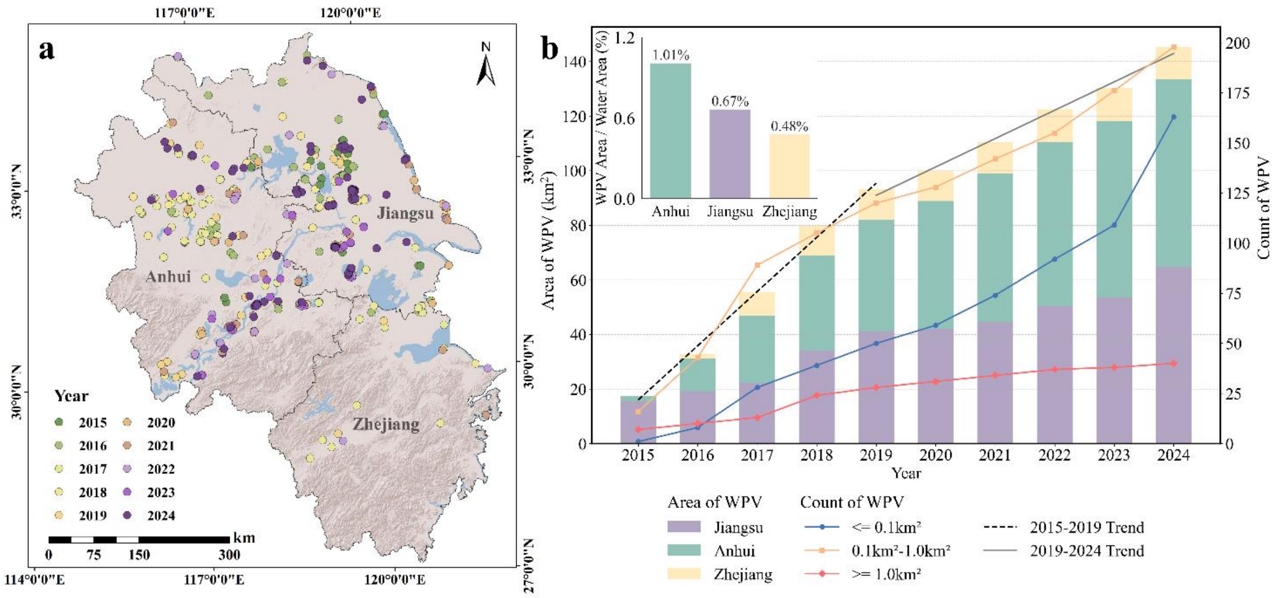


Figure 9. Spatiotemporal evolution of WPV installations from 2015 to 2024. a: Spatial distribution of WPV installations from 2015 to 2024 in Jiangsu, Anhui, and Zhejiang provinces, colored by year; b: Temporal evolution of WPV area and count during 2015–2024, where the stacked bars represent WPV area contributions from each province, and the lines indicate the count of WPV installations classified by size (≤0.1 km^2, 0.1–1.0 km^2, ≥1.0 km^2). Dashed and solid lines represent the fitting trends for 2015–2019 and 2019–2024, respectively. The inset bar chart (upper left) illustrates the proportion of WPV area relative to the water area within each province in 2024. (source: © 2014 Esri).

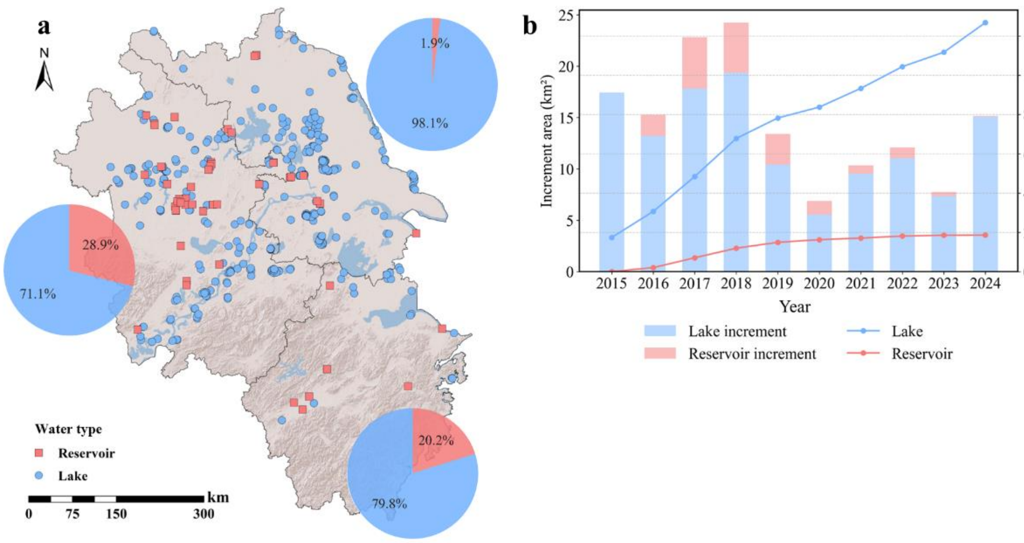


Figure 10. Spatial and temporal characteristics of WPV systems on different water body types in the Yangtze River Delta. a: Spatial distribution of WPVs on lakes (blue circles) and reservoirs (red squares), with pie charts indicating the proportion of WPV area on each water type in Jiangsu (top right), Zhejiang (bottom right), and Anhui (left); b: Annual WPV area increments on lakes and reservoirs from 2015 to 2024. Bars represent yearly increments, and lines represent cumulative area on each water type (source: © 2014 Esri).

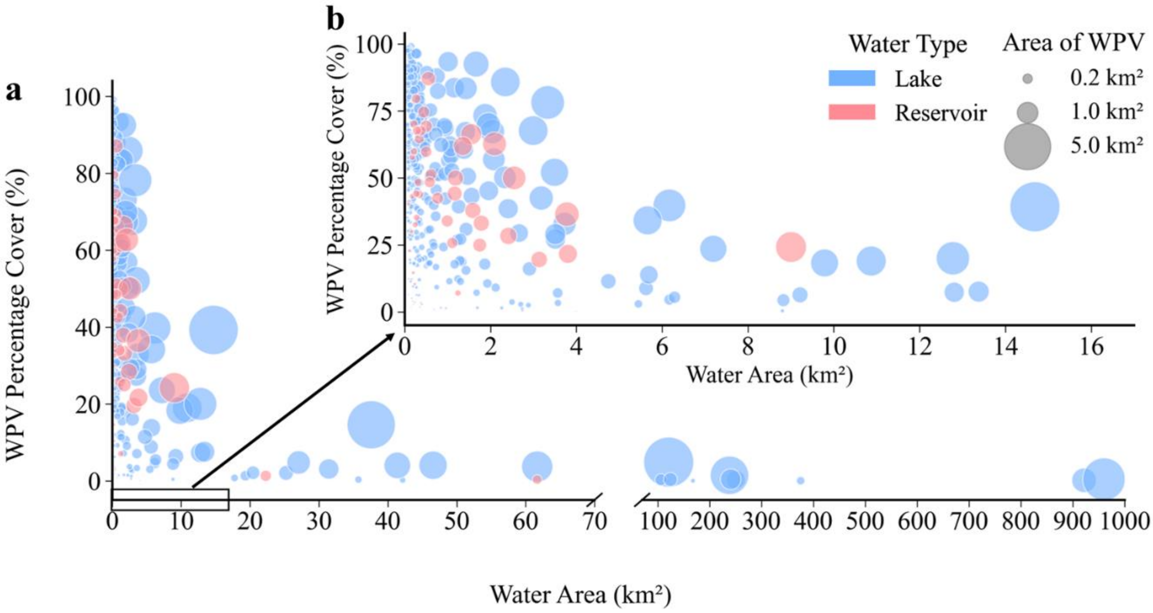


Figure 11. Relationship between water area and WPV percentage cover across lakes and reservoirs. a: Distribution of WPV percentage cover (%) against water body area (km^2), with circle size representing the WPV area; b: Enlarged view of the 0–10 km^2 water area range to highlight clustering patterns. Blue and red colors represent lakes and reservoirs, respectively, and circle size denotes WPV area.

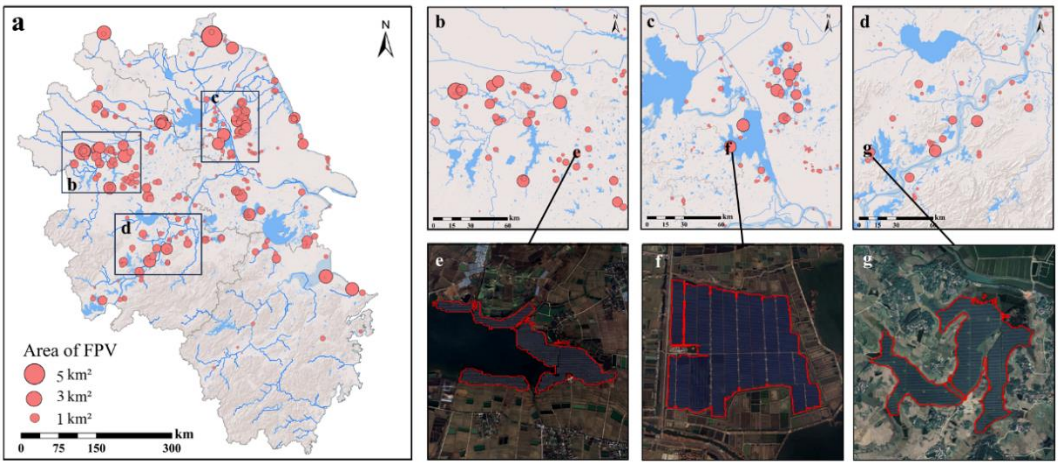


Figure 12. Spatial Clustering and Local Layouts of WPV Installations. a: Distribution of WPV systems across Jiangsu, Anhui, and Zhejiang provinces. The size of the red circles represents the area of each WPV system; b–d: Enlarged views of typical clustered regions highlighted in a; e–g: Satellite images showing detailed layouts of selected WPV systems corresponding to the locations indicated in b–d, with WPV boundaries outlined in red (source: Google Earth Pro, © 2025 Airbus, © 2025 Airbus / CNES, © 2025 Maxar Technologies, © 2014 Esri).

Table 1. Validation of the accuracy of the classification results and comparison with other datasets.

Dataset	A global inventory of PV	Global Renewables Watch - Solar	Our dataset
User Accuracy (%)-WPV	73.8	73.1	96.9
Producer Accuracy (%)-WPV	65.6	64.5	97.0
Overall Accuracy (%)	82.5	81.8	97.5
Kappa Coefficient	0.651	0.637	0.949

Table 2. The top 5 largest WPV areas in each province.

Province	City	Year	Water Type	Area (km^2)	Water Area (km^2)	Ratio (%)	Coordinates
	Fuyang	2022	Lake	3.53	4.21	83.91	32°48′N, 116°14′E
	Huainan	2017	Lake	2.29	61.67	3.72	32°41′N, 117°07′E
Anhui	Lu'an	2016	Lake	1.97	238.35	0.83	32°08′N, 116°46′E
	Chuzhou	2018	Lake	1.95	960.42	0.20	32°48′N, 119°07′E
	Anqing	2021	Lake	1.93	2.99	64.54	30°47′N, 117°31′E
	Lianyungang	2024	Lake	5.76	14.69	39.25	34°42′N, 119°09′E
	Xuzhou	2022	Lake	2.60	3.33	78.18	34°54′N, 116°50′E
Jiangsu	Suqian	2018	Lake	2.55	37.58	6.78	33°15′N, 117°57′E
	Yangzhou	2019	Lake	2.22	121.18	1.83	33°17′N, 119°41′E
	Huaian	2015	Lake	2.12	960.42	0.22	32°56′N, 119°14′E

	Ningbo	2017	Lake	2.40	6.17	38.88	30°17′N, 121°06′E
	Ningbo	2018	Reservoir	2.18	9.00	24.21	30°01′N, 121°37′E
Zhejiang	Jiaxing	2017	Lake	1.35	3.18	42.52	30°56′N, 120°48′E
	Huzhou	2017	Lake	0.96	3.51	27.23	30°40′N, 120°08′E
	Huzhou	2017	Reservoir	0.60	1.59	37.88	30°49′N, 119°43′E