

Article

A Comparison of Recent Global Time-Series Land Cover Products

Peilin Li ¹, Yan Wang ¹, Chisheng Wang ¹, Lin Tian ¹, Meijiao Lin ¹, Siyao Xu ² and Chuanhua Zhu ^{1,*}

¹ Ministry of Natural Resources (MNR) Key Laboratory for Geo-Environmental Monitoring of Great Bay Area & Guangdong Key Laboratory of Urban Informatics, School of Architecture & Urban Planning, Shenzhen University, Shenzhen 518060, China; 2021303009@email.szu.edu.cn (P.L.); 2022104003@email.szu.edu.cn (Y.W.); wangchisheng@szu.edu.cn (C.W.); tianlin@szu.edu.cn (L.T.); 2021071101@email.szu.edu.cn (M.L.)

² College of Electronics and Information Engineering, Shenzhen University, Shenzhen 518060, China; 2021280263@email.szu.edu.cn

* Correspondence: czhu@szu.edu.cn

Abstract: Accurate and reliable land cover data are essential for environmental monitoring, climate research, and sustainable land management. However, the proliferation of multi-source global land cover datasets with long time series poses challenges for selecting the best products for specific applications. Existing assessments often lack systematic comparisons of classification accuracy and time consistency across geographic areas. This study addresses the critical gap in cross-product comparability by systematically comparing five recent global time-series land cover products (GLC_FCS30D, Esri Land Cover, MCD12Q1, ESA CCI, and Dynamic World) against a reference dataset (CGLS-LC100). Through a unified classification system, resolution resampling, and random sampling validation, we assessed their classification accuracy and time-series change accuracy across three transitional regions representing diverse environmental contexts: rapidly urbanizing regions, agriculturally intensive zones, and high-latitude forested areas. The results indicate that while datasets exhibit spatial consistency, significant discrepancies exist in land cover classification, with each dataset demonstrating varying levels of accuracy depending on the environmental context and land cover type. High-resolution products (e.g., GLC_FCS30D, Dynamic World) are optimal for monitoring fragmented landscapes and urban expansion, whereas long-term datasets (e.g., ESA CCI, MCD12Q1) suit climate trend analysis in stable ecosystems. Based on the evaluation, we provide generalized guidance for dataset selection aligned with land cover types and monitoring objectives, emphasizing the need for region-specific and application-oriented choices. This study highlights challenges in dynamic datasets, including classification system discrepancies, resolution effects, and reference data limitations, and suggests that future advancements should focus on improving classification algorithms, refining sampling methods, and developing assessment systems that incorporate high-precision, real-time validation data.

Keywords: land cover; time series; product comparison; land cover change; tracking accuracy



Academic Editor: Georgios Mallinis

Received: 17 February 2025

Revised: 8 April 2025

Accepted: 14 April 2025

Published: 16 April 2025

Citation: Li, P.; Wang, Y.; Wang, C.; Tian, L.; Lin, M.; Xu, S.; Zhu, C. A Comparison of Recent Global Time-Series Land Cover Products. *Remote Sens.* **2025**, *17*, 1417. <https://doi.org/10.3390/rs17081417>

Copyright: © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Land cover survey and mapping is an important basis for optimizing agricultural management [1], improving the accuracy of forest cover estimation [2], and promoting sustainable development [3]. Current land cover products are mainly divided into two technological paradigms: The first is single-year high-precision reference datasets (e.g., GLC_FCS30D), which realize subpixel-level classification through multi-source data fusion.

Although these do not explicitly construct time series, multi-baseline-year comparison can indirectly support dynamic land cover analysis. The second category is local temporal dynamic datasets, which use standardized algorithms to generate continuous temporal products. Examples include the European Space Agency Climate Change Initiative (ESA CCI), which has provided annual composite data at 300 m resolution since 1992, and Dynamic World, which updates 10 m resolution images every 5 days through the Long Short-Term Memory (LSTM) model. These two kinds of technological breakthroughs make dynamic land cover analysis possible and profoundly reshape the theoretical and methodological system of land change research.

Dynamic land cover data, with the advantages of multi-scale and high-frequency monitoring, are driving the enhancement of management efficiency across various fields [4,5]. In recent years, with the continuous advancement of remote sensing technology and the diversification of data acquisition methods, the number and types of global time-series land cover datasets have shown a rapid growth trend. These datasets originate from different satellite sensors, research institutions, and projects, and they vary in spatial resolution, temporal resolution, and classification systems. In addition to the aforementioned datasets, such as GLC_FCS30D, ESA CCI, and Dynamic World, there are also products like MODIS MCD12Q1 and Esri Land Cover, each with unique data characteristics and application scenarios. Many global time-series land cover datasets now span long time periods, providing rich data support for studying long-term land cover changes. For instance, the ESA CCI has been offering annual composite land cover data at 300 m resolution since 1992. Moreover, datasets like Dynamic World make it possible to conduct real-time monitoring and short-term change analysis through more frequent update cycles, such as updating 10 m resolution imagery every five days. However, faced with the multitude of global time-series land cover datasets, users often find themselves perplexed in practical applications. Differences in classification accuracy, temporal consistency, and spatial resolution among datasets make it difficult for users to determine which dataset is most suitable for their specific research or application needs.

There are several deficiencies in the current evaluation of global time-series land cover datasets, which are primarily manifested in the following aspects:

- (1) Lack of systematic comparison. At present, there is a relative scarcity of systematic comparisons regarding the classification accuracy and temporal consistency of different land cover datasets, along with a lack of unified evaluation standards and methods. Different studies often employ varying evaluation metrics and data samples, resulting in outcomes that are difficult to compare and integrate. Moreover, there is a lack of systematic comparison between geographical regions, with most research focusing on specific regions or the global scale, neglecting the differences and connections between regions. For instance, Yang et al. [6] proposed to assess the reliability of time-series land cover products through Hidden Markov Model (HMM) joint probability, combining classification performance with spatiotemporal relationships to validate the land cover data in the Poyang Lake Ecological and Economic Zone. Narumasa et al. [7] developed a spatiotemporal accuracy assessment method based on Geographically Weighted Logistic Regression (Logistic GWR) for the rapidly urbanizing Jakarta Metropolitan Area using MODIS time-series data from 2001 to 2013. While this regionalized validation framework can reveal local accuracy heterogeneity, it relies on a single dataset (MODIS) and lacks an established cross-resolution alignment protocol, preventing the support of spatiotemporal comparability analysis across multiple products. Therefore, a global-scale evaluation of the temporal consistency of dynamic land cover datasets remains lacking.

- (2) Insufficiency in dynamic assessment: Despite the increasing number of dynamic land cover datasets, the systematic evaluation of their dynamic characteristics, particularly in terms of temporal consistency, is lagging. Current assessments primarily focus on static classification accuracy. For instance, Herold et al. [8] conducted an analysis of spatial consistency and uncertainty for global-scale land cover products, revealing the limitations of current 1 km resolution products in complex landscapes by harmonizing classification standards across multiple datasets. Tsendbazar et al. [9] systematically assessed the strengths and limitations of existing datasets in supporting various application scenarios from a user's perspective. However, many studies overlook the coherence and stability of datasets over time, which are crucial for monitoring land cover changes and predicting future trends. Currently, there is a lack of an effective validation framework to assess this key attribute. For example, ensuring that the land cover classification results of long-time-series datasets are consistent and comparable over time is an urgent issue that needs to be addressed.
- (3) Deficiencies in technical research: Existing studies still fall short in addressing issues such as scale dependency differences, uncertainty quantification, and the fusion of multi-source datasets.

Firstly, the impact of scale effects on data accuracy shows a cascading amplification characteristic between coarse-resolution and high-resolution products. Moody et al. [10] systematically revealed the cascading effects of spatial-resolution differences on land cover classification accuracy, noting that coarse-resolution datasets may significantly underestimate the proportion of rare land classes in fragmented landscapes, such as wetlands and urban green spaces. Tudesque et al. [11] found evidence in the Adour-Garonne river basin that errors in land cover classification, which vary with scale, not only occur in the classification, but also greatly harm subsequent applications by misaligning how land classes respond to the environment, highlighting the need for a validation method that ensures spatial consistency across different resolutions for the accuracy of ecological models. There is a lack of effective methods for comparing and fusing datasets of different resolutions, with coarse- and fine-resolution products differing in information expression and application requirements. How to achieve accurate conversion and integration across scales remains an area in need of further research.

Secondly, the quantification of uncertainty in the dynamics of time-series land cover is underrepresented in existing studies, which often focus on the classification accuracy of static data. For example, Waśniewski et al. [12] improved the classification accuracy of Sentinel-2 data through optimized sample selection and DEM fusion, but the validation relied solely on single-phase confusion matrices (such as Kappa coefficients), without assessing the logical rationality of inter-annual changes. Abercrombie et al. [13] significantly reduced the magnitude of spurious inter-annual changes in coarse-resolution (300–500 m) land cover products by constraining label transition probabilities with HHM. However, such methods are still limited to internal time-smoothing optimizations within a single dataset and do not address the issue of cross-resolution comparability between multi-source products, leading to challenges in ensuring the compatibility and consistency of different-resolution data in land cover dynamics monitoring. Existing research predominantly focuses on statistical descriptions of classification accuracy, with less attention given to uncertainty in the temporal dimension, which lacks a systematic validation framework to comprehensively assess data reliability.

Finally, research on the fusion of multi-source datasets is relatively weak. To improve the cross-comparability of multi-source land cover products, the academic community has investigated harmonization techniques for multi-source land cover datasets, with classification system calibration and resolution resampling trade-offs being key aspects

of this process [14]. In terms of classification system calibration, Liu et al. [15] released the GLASS-GLC global long-time-series land cover product, which improved the annual average classification accuracy to 82.81% by constructing a full-season universal sample library and temporal–spatial consistency post-processing techniques. In terms of resolution coordination, when fusing Sentinel-2 and Landsat-8 data, they found that bilinear interpolation led to a decrease in user accuracy of 8–12% in farmland boundary representation and proposed a sub-pixel algorithm based on spectral unmixing to mitigate mixed-pixel errors. However, such methods are often limited to single sensors or land classes and have not yet been systematically integrated into a unified evaluation framework.

There is an urgent need to construct a framework that coordinates the interoperability of classification systems with the optimization of dynamic resolutions. Herold et al. [8] proposed a semantic mapping theory across classification systems based on the FAO's LCCS framework, but it did not address the scale-effect issue when migrating from coarse-resolution products (such as the 5 km of GLASS-GLC) to high resolutions. Meanwhile, Fritz et al. [16] reduced misclassification rates of mixed land types through an ontological dynamic mapping tool, yet they did not integrate the impact of resolution resampling on classification logic. These technical bottlenecks and the gap between application needs highlight the urgency of systematically evaluating the spatiotemporal consistency of multi-source time-series land cover datasets.

The current global time-series land cover datasets suffer from deficiencies in systematic comparison, dynamic assessment, and technical research. This study addresses these issues by conducting a comprehensive analysis of five recent global time-series land cover products (GLC_FCS30D, Esri Land Cover, MCD12Q1, ESA CCI, and Dynamic World), with CGLS-LC100 serving as a reference dataset. A systematic evaluation framework was constructed and tested in three case study regions (the Guangdong–Hong Kong–Macao Greater Bay Area, the Visalia region in the United States, and the Norway–Sweden border area). This framework employs methodological approaches, including the unification of classification systems, resolution resampling, and random sampling validation to realize the differences, strengths, and limitations of multi-source time-series land cover datasets in terms of classification accuracy and temporal sequence change accuracy. This fills the existing gap in the literature and provides a scientific basis for selecting the optimal dataset for different regions and applications.

2. Datasets and Study Area

The selection of comparative and validation datasets in this study was guided by four criteria to ensure methodological rigor: (1) temporal continuity covering major global change periods (spanning multi-decadal continuity to near-real-time updates), (2) representative spatial resolutions spanning 10 m–500 m, (3) compatibility with the FAO LCCS classification framework, and (4) open-access availability through authoritative platforms. These criteria enable the systematic evaluation of time-series land cover products across sensor types, algorithm paradigms, and spatial–temporal scales. To comprehensively assess the performance of existing long-term time-series land cover products, this study selected five publicly available global time-series land cover products (GLC_FCS30D, Esri Land Cover, MCD12Q1, ESA CCI, and Dynamic World) as comparative datasets and conducted a comparative analysis with one validation dataset, CGLS-LC100.

2.1. Datasets

2.1.1. Comparative Datasets

Five recent global time-series land cover products were selected for comparison: (1) GLC-FCS30D (30 m resolution) is the first global land cover dynamics monitoring

product with a fine classification system, derived from dense-time-series Landsat imagery and continuous change detection, covering 1985–2022 [17]; (2) Esri Land Cover (10 m resolution) employs deep learning on Sentinel-2 imagery with annual updates since 2017, trained on billions of labeled pixels from National Geographic Society [18]; (3) MCD12Q1 (500 m resolution) has provided annual MODIS-derived classifications since 2001 and is widely recognized in ecological and climate studies [19]; (4) ESA CCI (300 m resolution) delivers long-term change detection (1992–2024), validated with 74.72% accuracy in tropical basins [18,20,21]; and (5) Dynamic World (10 m resolution) features near-real-time Sentinel-2 processing through Google’s AI platform, offering biome-specific classifications since 2015 [22].

2.1.2. Validation Dataset

CGLS-LC100 (100 m resolution) from Copernicus Global Land Service was adopted for validation, providing continuous field layers and discrete classifications from 2015 to 2019. Derived from PROBA-V time series and high-quality training sites, this product achieves 80% Level-1 accuracy with planned Sentinel-based annual updates from 2020 [23]. Its selection balances resolution adaptability (superior to traditional 500 m products) and cross-product compatibility, while serving as a recognized benchmark in global validation studies.

2.1.3. Product Characteristic

This section analyzes the core characteristics of the products across temporal, spatial, and categorical dimensions (Table 1) to assess their usability. Some datasets are available for free, while commercial datasets require purchase, which limits their broader applicability. High-resolution data, while offering more precise information, are large in size and computationally intensive. In contrast, low-resolution data, although lacking in detail, are well suited for large-scale studies due to their smaller size and ease of processing. High-resolution products excel in urban boundary identification and the classification of fragmented land cover patterns, while long-time-series and wide-coverage datasets are more appropriate for macro-trend analysis.

Table 1. Main parameters of each dataset.

Product Name	Abbreviation	Temporal Characteristics		Spatial Characteristics		Criteria for Classification	
		Year of Coverage	Calculation Period	Resolution	Scope of Coverage	Classification Method	Land Cover Categories
CGLS-LC100	CGLS	2015–2019	1 year	100	global	Supervised Classification (Random Forest, Decision Tree, etc.)	10
GLC_FCS30D	FCS	2015–2021	1 year	30			30
Esri Land Cover	Esri	2015–2024	1 year	10			10
MCD12Q1	MCD	2001–2024	1 year	500			17
ESA CCI	CCI	1992–2020	1 year	300			22
Dynamic World	DW	2015–2024	2–5 days	10		Convolutional Neural Network (FCNN)	9

2.2. Validation Area

Given that transition zones are rich in environmental complexity, economic diversity, and significant land cover changes [24], and have been selected as study objects in numerous studies [25,26], we chose the following three transitional study areas for evaluating our data products (Figure 1): (a) Visalia in the United States (2210.43 km²), characterized by a land cover pattern predominantly influenced by agriculture, is significantly shaped by natural factors such as mountainous topography, water availability, and soil conditions, as well as agricultural policies that dictate crop types and farming practices. (b) The central part of the

Norwegian–Swedish border (18,999.96 km²) is situated in a high-latitude area, dominated by forest vegetation, snow, and grasslands; land cover is significantly constrained by climatic factors. (c) The central part of the Guangdong–Hong Kong–Macao Greater Bay Area in China (GHMA, 10,216.49 km²). As one of the most rapidly developing regions in the world, the GHMA is marked by a high concentration of urbanization and dramatic land cover changes. The area is dominated by built-up urban areas that have expanded rapidly in recent decades, alongside remnant forested lands. This region exemplifies the challenges and complexities of land cover mapping in densely populated and economically dynamic areas, where the pace of change is fast, and the types of land cover transitions are diverse.

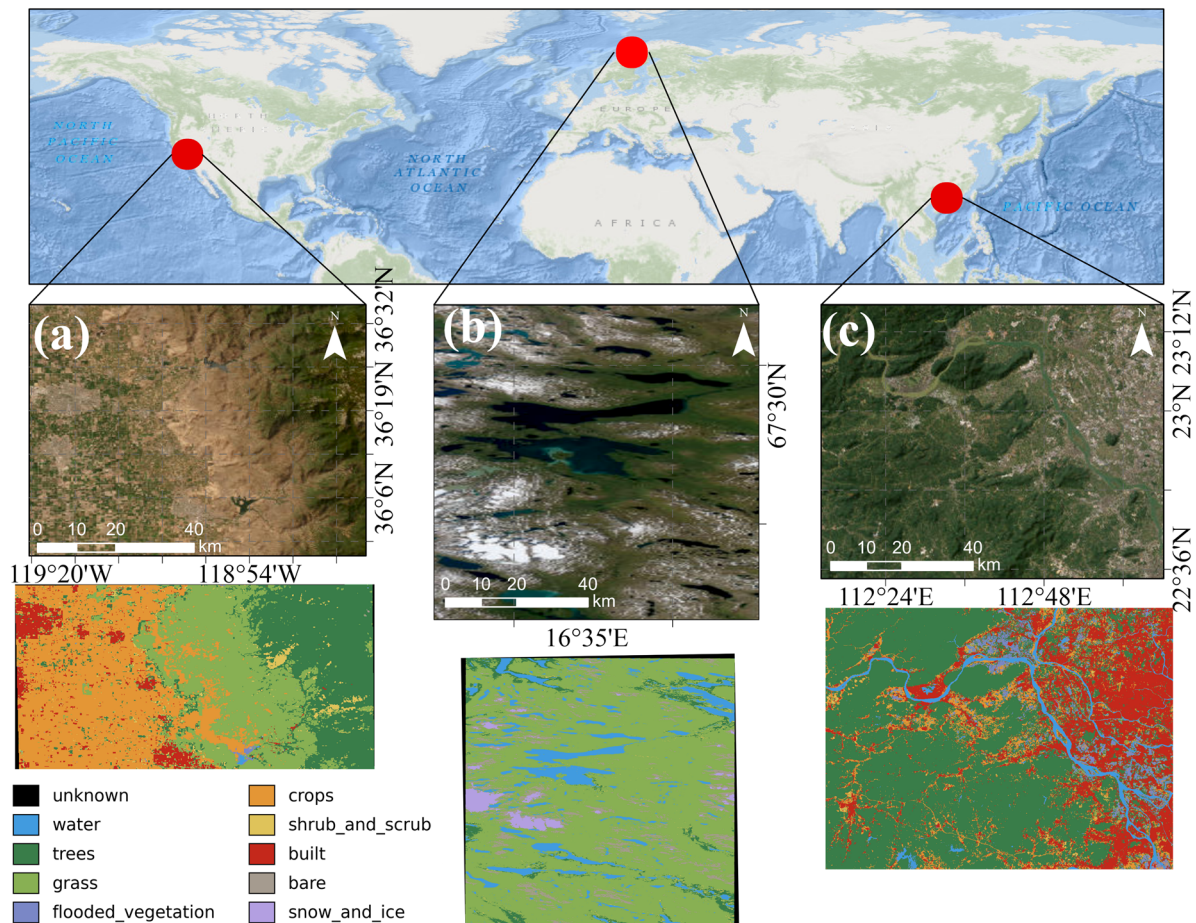


Figure 1. Schematic diagram of the study area. (a) Visalia, USA; (b) Central Norwegian–Swedish boundary; (c) Central Guangdong–Hong Kong–Macao Greater Bay Area.

Collectively, these regions span a gradient of environmental conditions from high-latitude forests to urbanized coastal areas and represent different stages of economic development and land cover transition intensities. They allow for a robust assessment of data product performance across these diverse contexts, ensuring that our evaluation captures the variability and complexity of global land cover dynamics.

3. Methods

3.1. Preparation for Product Comparison

3.1.1. Classification System Unification

This study utilizes the Land Cover Classification System (LCCS) developed by the Food and Agriculture Organization of the United Nations. It is a standardized and versatile

system widely used for global land cover mapping and analysis [22]. Its main types are divided into eight categories (Table 2).

Table 2. System framework of LCCS.

Dichotomous Phase		
Primarily Vegetated	Terrestrial	Cultivated and Managed Terrestrial Areas
		Natural and Semi-Natural Terrestrial Vegetation
	Aquatic or Regularly Flooded	Cultivated Aquatic or Regularly Flooded Areas
		Natural and Semi-Natural Aquatic or Regularly Flooded Vegetation
Primarily Non-Vegetated	Terrestrial	Artificial Surfaces and Associated Areas
		Bare Areas
	Aquatic or Regularly Flooded	Artificial Waterbodies, Snow and Ice
		Natural Waterbodies, Snow and Ice

To create a unified framework for comparing multi-source data, this study adopts a three-level mapping strategy: (1) Direct mapping: Common categories are directly matched based on consistent definitions. (2) Feature-associated mapping: For categories that differ, spectral features are combined with geographic knowledge, and multi-temporal data are used to assist in categorization [27]. For example, in the FCS product, salt marsh and tidal flat are mapped to the herbaceous wetland category in the CGLS product based on their shared wetland features. (3) Expert validation: For fuzzy categories, field validation and expert consultation are used to determine appropriate attribution, minimizing subjective bias.

Based on the above strategy, we reclassified the various land classification products according to their attributes and definitions and obtained the following results (Table 3):

Table 3. Reclassification.

Reclassified Category	CGLS (Validation Dataset)	FCS	CCI	MCD	Esri	DW
Unknown (0)	Unknown (0)	No data (0), filled value (250)	No data (0)	Unclassified (255)	No data (1), clouds (11)	
Water (1)	Permanent water bodies (80), oceans, seas (200)	Swamp (181), marsh (182), flooded flat (183), water body (210)	Water bodies (210)	Water bodies (17)	Water (2)	Water (0)
Trees (2)	Closed forest, evergreen needle leaf (111), evergreen broad leaf (112), deciduous needle leaf (113), deciduous broad leaf (114), ... mixed (115), not matching any of the other definitions (116), evergreen needle leaf (121), evergreen broad leaf (122), deciduous needle leaf (123), deciduous broad leaf (124), mixed (125), open forest, not matching any of the other definitions (126)	Tree cover (12), open evergreen broadleaved forest (51), closed ... (52), open deciduous broadleaved forest (61), closed ... (62), open evergreen needle-leaved forest (71), closed ... (72), open deciduous needle-leaved forest (81), closed ... (82), open mixed leaf forest (91), closed mixed leaf forest (92)	Tree or shrub cover (12), mosaic natural vegetation (40), tree cover, broadleaved, evergreen, closed to open (50), ... closed (60), ... open (61), T. needleleaved, evergreen, closed (70), ... open (71), T. needleleaved, deciduous, closed (80), ... open (81), T. mixed leaf type (90), mosaic tree and shrub (100), mosaic herbaceous cover (110), sparse vegetation (150), sparse tree (151)	Evergreen needleleaf forests (1), evergreen broadleaf forests (2), deciduous needleleaf forests (3), deciduous broadleaf forests (4), mixed forests (5), savannas (9)	Trees (3)	Trees (1)
Grass (3)	Herbaceous vegetation (30)	Grassland (130), sparse herbaceous (153)	Grassland (130), herbaceous cover (11), sparse herbaceous cover (153)	Grasslands (10)	Grass (4)	Grass (2)
Flooded_vegetation (4)	Herbaceous wetland (90)	Salt marsh (186), tidal flat (187)	Flooded, fresh or brackish water (160), flooded, saline water (170), fresh/saline/brackish water (180)	Permanent wetlands (11)	Flooded vegetation (5)	Flooded_vegetation (3)

Table 3. Cont.

Reclassified Category	CGLS (Validation Dataset)	FCS	CCI	MCD	Esri	DW
Crops (5)	Cultivated and managed vegetation (40)	Rainfed cropland (10), herbaceous cover cropland (11), irrigated cropland (20)	Crop, rainfed (10), crop, irrigated or post-flooding (20)	Croplands (12), natural vegetation mosaics (14)	Crops (6)	Crops (4)
Shrub and scrub (6)	Shrubs (20)	Shrubland (120), evergreen shrubland (121), deciduous shrubland (122), sparse shrubland (152), sparse herbaceous (153), mangrove (185)	Shrubland (120), evergreen shrubland (121), deciduous shrubland (122), sparse vegetation (150), sparse tree (151), sparse shrub (152), sparse herbaceous cover (153)	Closed shrublands (6), open shrublands (7), woody savannas (8)	Scrub (7)	Shrub_and_scrub (5)
Built (7)	Urban/built up (50)	Impervious surfaces (190)	Urban areas (190)	Urban and built-up lands (13)	Built (8)	Built (6)
Bare (8)	Moss and lichen (100), bare vegetation (60)	Lichens and mosses (140), sparse vegetation (150), saline (184), bare areas (200), consolidated bare areas (201), unconsolidated bare areas (202)	Lichens and mosses (140), bare areas (200), consolidated bare areas (201), unconsolidated bare areas (202)	Barren (16)	Bare ground (9)	Bare (7)
Snow and ice (9)	Snow and ice (70)	Permanent ice and snow (220)	Permanent snow and ice (220)	Permanent snow and ice (15)	Snow/ice (10)	Snow_and_ice (8)

3.1.2. Resampling

In the analysis of the multi-source dataset, the significant resolution disparity between products presents a major challenge, as it can introduce bias and uncertainty in accuracy assessments. To address this issue, all products in this study were resampled to a 100 m resolution (consistent with the CGLS products) to eliminate the effects of resolution differences and ensure a fair comparison of each product at the same scale.

For high-resolution datasets (e.g., 10 m), resampling to 100 m was implemented through pixel aggregation using the nearest neighbor method. This approach avoids averaging categorical land cover classes and preserves the dominant class at the native resolution.

For coarse-resolution datasets (e.g., 500 m MODIS), resampling to 100 m required pixel disaggregation. Each 500 m pixel was divided into a 5×5 grid of 100 m cells. The value of the original 500 m pixel was replicated across all 25 subdivided 100 m cells. This method ensures geometric alignment with other 100 m products without introducing interpolated values that could lead to unrealistic land cover classifications.

In conclusion, the nearest neighbor method was used for upsampling from 10 m to 100 m to maintain the categorical consistency of land cover labels. For downscaling from 500 m to 100 m, a replication method was used to avoid interpolation artifacts that could distort accuracy assessments.

3.2. Assessment Criteria

In remote sensing land cover classification and time-series change analysis, accuracy assessment is crucial for evaluating the reliability of the results. This study employs both classification accuracy and time-series accuracy for a comprehensive evaluation to ensure the reliability of the research findings.

3.2.1. Classification Accuracy

The assessment period for all validation samples spans from 2015 to 2019. In this study, the confusion matrix was used to analyze the classification results, and both the Kappa coefficient and Overall Accuracy (OA) were calculated. OA reflects the proportion of correctly classified pixels, while the Kappa coefficient accounts for the agreement between the classification results and random chance, thereby reducing the risk of misleading accuracy metrics caused by class imbalances. This combination enhances the comprehensiveness and reliability of the results. To further mitigate mapping errors, a 5×5 pixel window (500×500 m) applied to the resampled 100 m resolution dataset was implemented as the minimum statistical unit for sampling. The predominant land cover type within each window was determined through a hierarchical decision rule: (1) selection of the class with the highest pixel proportion; (2) prioritization of spatially continuous classes from adjacent windows in cases of tied proportions; (3) exclusion of unresolved cases labeled as mixed class from accuracy calculations to ensure reliability by focusing on areas with unambiguous dominant land cover classes. To ensure robustness, 50% of the total pixels were randomly selected from the entire image as validation points in each independent iteration (three iterations in total). This approach addressed three critical considerations: (1) partial sampling avoided spatial autocorrelation bias by disrupting pixel continuity and reduced computational costs; (2) averaging results across three independent iterations minimized random errors and enhanced statistical robustness; (3) the 50% threshold balanced precision and efficiency for large-scale 100 m resolution data. Although 50% of pixels were excluded in each iteration, overlapping subsets across three independent trials ensured comprehensive coverage of the entire image statistically. The final OA was derived from the average of these iterations, effectively balancing spatial detail preservation with statistical

reliability. No additional validation steps are required, as the methodology inherently accounts for the full dataset through repeated sampling.

Kappa coefficient formula:

$$K = \frac{p_0 - p_e}{1 - p_e}, \quad (1)$$

where p_0 denotes the actual observed classification accuracy, equivalent to OA, and p_e denotes the expected value of the random classification accuracy.

p_0 formula:

$$p_0 = \frac{\sum_{i=1}^n N_{ii}}{N}, \quad (2)$$

p_e formula:

$$p_e = \frac{\sum_{i=1}^n (N_{i+} \cdot N_{+i})}{N^2}, \quad (3)$$

OA formula:

$$OA = \frac{\sum_{i=1}^n N_{ii}}{N}, \quad (4)$$

where N_{ii} is the number of correctly categorized samples for class i , and N is the total number of validation samples in each iteration (50% of total pixels, excluding unresolved mixed-class cases). The final OA is the longitudinal average of three independent iterations, each using a randomly selected 50% subset of pixels.

Producer Accuracy (PA) formula:

$$PA_i = \frac{N_{ii}}{N_{i+}}, \quad (5)$$

where N_{i+} denotes the total number of reference samples in class i .

The Average Change metric quantifies interannual variations in classification consistency, calculated as the mean annual difference in PA values across consecutive years. For each land cover category and product combination, the year-to-year PA differences are computed as follows:

$$\Delta PA_t = PA_t - PA_{t-1}$$

where ΔPA_t represents the change in Producer Accuracy from year $t - 1$ to year t .

These annual differences are then averaged across all available year pairs to derive the Average Change:

$$Average\ Change = \frac{1}{n} \sum_{t=1}^n \Delta PA_t$$

where n represents the number of valid year intervals.

This metric reflects the directional stability of classification performance, with positive values indicating systematic PA improvements and negative values suggesting accuracy deterioration over time. The temporal resolution of this analysis was maintained at annual intervals to capture nuanced interannual dynamics while mitigating seasonal variability.

3.2.2. Temporal Accuracy

Temporal accuracy assesses the consistency of land cover types over time by comparing classification results with ground reference data. A year-by-year comparison with time spans of 1–5 years was used to check the consistency of land cover changes at sampling sites with validation data.

To capture both short- and long-term trends, a moving window approach was adopted to calculate accuracy across all consecutive year-pairs from 2015 to 2019. The assessment steps were as follows:

- (1) Randomly select 20% of sample sites and record year-to-year land cover changes. This sampling ratio was determined to balance computational efficiency with statistical representativeness while minimizing spatial autocorrelation through random selection.
- (2) For each time span T (1–5 years), calculate the correct-match percentage by comparing classification and validation data across all T -year intervals.
- (3) Average the accuracy across all T -year intervals to reduce temporal bias.
- (4) Repeat the calculation three times and take the average as the final accuracy.

The temporal accuracy is calculated using the following formula:

$$A_T = \frac{\sum_{i=1}^n M_i}{N}, \quad (6)$$

where T is the time span; A_T is the accuracy for T years; M_i indicates the full temporal consistency of the sampling point (1 if all years in the T -year interval match validation data; 0 otherwise); and N is the total number of sampling points.

4. Results

This study presents an initial analysis of land cover changes by systematically organizing land cover data across three regions over time.

The central portion of the GHMA is distinguished by its high urbanization rate and diverse land cover types (Figure 2). Between 2015 and 2019, the central urban area experienced significant expansion, with large tracts of cropland and low-density vegetation being replaced by urban land.

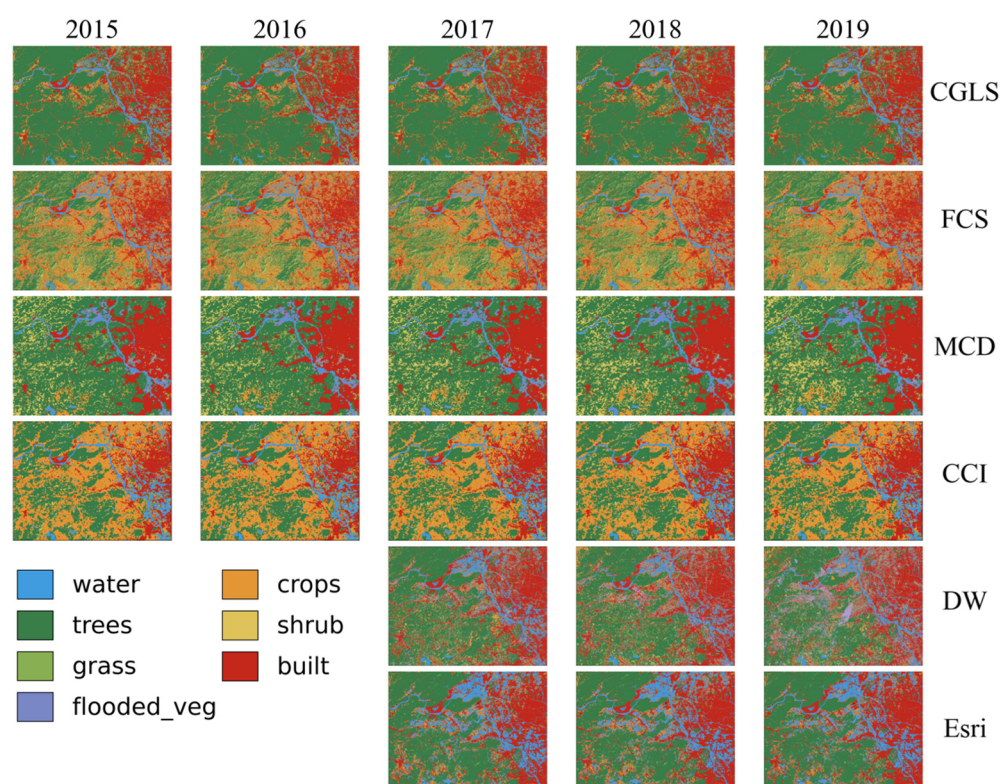


Figure 2. Overview of land cover changes in the central part of the Guangdong–Hong Kong–Macao Greater Bay Area.

In the Visalia region of the United States, the landscape is predominantly agrarian, characterized by vast farmlands (Figure 3).

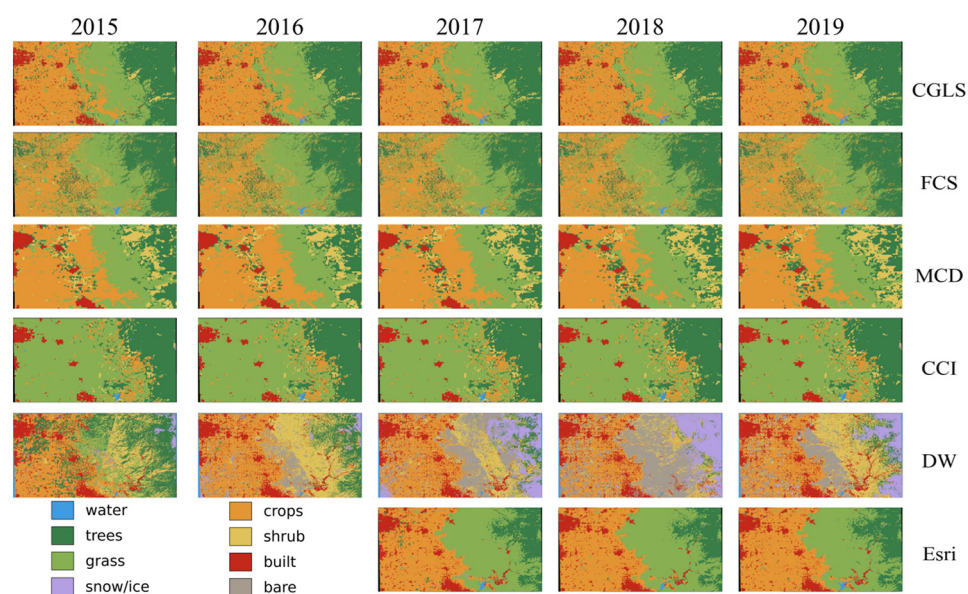


Figure 3. General overview of land cover change in Visalia, USA.

Between 2015 and 2019, the area of farmland remained relatively stable and also showed a downward trend, reflecting changes in agricultural practices and land management policies.

The central part of the Norwegian–Swedish border is located in the high latitudes of the northern hemisphere and is predominantly covered by forests (dark green), with scattered grasslands (light green) and shrubs (yellow), resulting in a high overall vegetation cover (Figure 4).

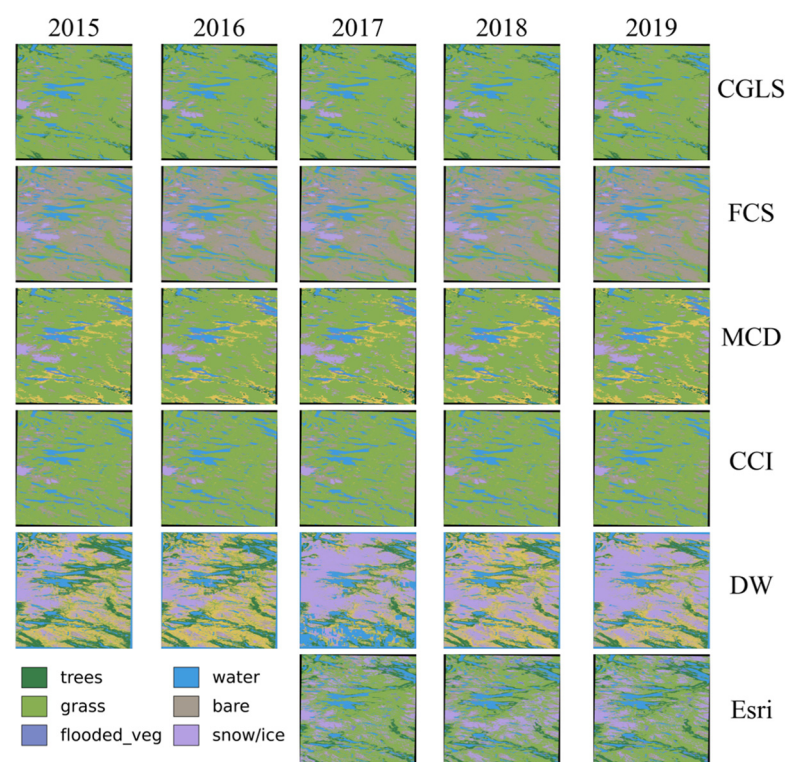


Figure 4. Overview of land cover change in the central part of the Norwegian–Swedish border.

Between 2015 and 2019, forest area remained stable, while grassland and shrub areas increased in some mountainous regions, reflecting the effects of climate warming and the natural recovery of vegetation.

Regional land cover exhibits spatial consistency but also significant differences across datasets. The CGLS dataset demonstrates consistent performance with minimal variation, particularly in forest and water body classifications. In contrast, the FCS and MCD datasets show greater fluctuations in arable and grassland classifications. The CCI dataset struggles to clearly distinguish between shrubs and grasses, leading to overlapping classifications in some regions. MCD is more accurate in classifying urban areas and bare land. The DW dataset stands out for its refined classification capabilities, especially in urban and bare land identification.

The classification of floodplain vegetation near water bodies varies across products. CGLS and Esri show greater sensitivity to floodplain vegetation, while FCS and MCD tend to misclassify portions of the floodplain as grassland or shrubs. For built-up areas, DW demonstrates a notable increase between 2017 and 2019, capturing urban sprawl with higher precision. These findings highlight the importance of selecting appropriate datasets based on specific research objectives and regional characteristics.

4.1. Classification Accuracy by Year

The classification accuracy of various products in the GHMA exhibited notable temporal variability between 2015 and 2019 (Figure 5). The MCD products demonstrated higher accuracy in the initial years but experienced a decline in subsequent years. In contrast, Esri products showed significant fluctuations, with an initial rise in accuracy followed by a subsequent decrease. The DW products exhibited higher accuracy exclusively in 2017, with relatively lower accuracy in other years. On the other hand, FCS and CCI products maintained stable accuracy levels, achieving moderate overall performance throughout the study period. The Kappa coefficient exhibited a similar trend to the OA, underscoring the strong classification consistency and reliability of these products.

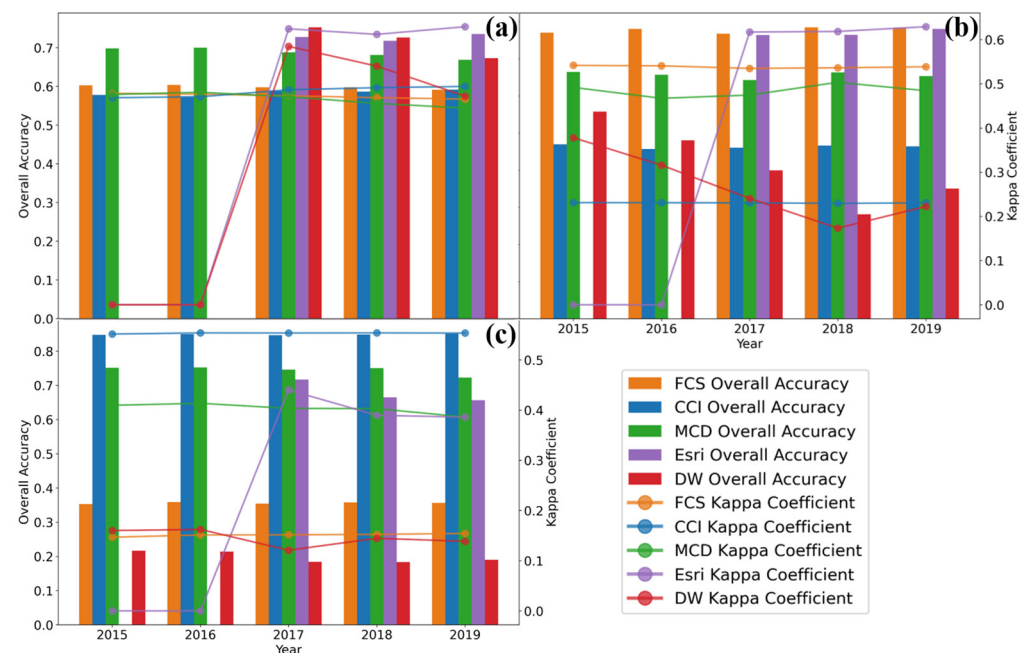


Figure 5. OA and Kappa coefficient. (a) Guangdong–Hong Kong–Macao Greater Bay Area; (b) Visalia, USA; (c) central Norway–Sweden junction.

In Visalia, the accuracy of the products varied significantly across the study period, with certain products performing well in specific years while displaying large annual fluctuations. The DW product showed higher accuracy in certain years, while the FCS and CCI products were stable, although they did not exceed the peaks of other products. The fluctuating Kappa coefficients reflect the complexity of the land cover types in the region and the limited environmental adaptability of the products.

For the central part of the Norway–Sweden border, the accuracy of the products remained relatively stable without significant fluctuations. While some products exhibited higher or lower accuracy, the OA levels were consistent across the study period. The Kappa coefficients showed smaller variations, indicating that classification accuracy and consistency were more stable, though there was limited potential for performance improvement, likely constrained by local environmental factors.

The analysis highlights the varying performance of different classification products across regions and years. While some products demonstrated higher accuracy in specific areas, others maintained consistent performance. The Kappa coefficients further corroborate these findings, emphasizing the importance of selecting appropriate products based on regional characteristics and research objectives.

In terms of product performance, the accuracy of FCS is stable in both the GHMA and Visalia, but generally moderate overall. For the central Norway–Sweden border, the accuracy of FCS is influenced by topographic and climatic factors [28,29]. The accuracy of CCI remains relatively stable at a medium level, although the specific value varies by region. MCD exhibits higher accuracy in the early years, followed by a decline in the Greater Bay Area, while its performance in Visalia and the central Norway–Swedish border differs. In these regions, MCD's accuracy shows smaller variations, reflecting its adaptability to environmental and land cover changes. The Esri product demonstrates more pronounced fluctuations in the Greater Bay Area, while its accuracy is more stable in the other two regions. The DW product shows exceptional performance in the GHMA and Visalia in certain years, but no such pattern is observed for the central part of the Norwegian–Swedish border.

4.2. Comparative Analysis of Different LULC Categories

This subsection compares the trends and causes behind the classification accuracy results of each product, based on the PA for each LULC category in the study area.

In the core region of the GHMA, where urban landscapes are interwoven with natural woodlands and complex topography, the PA of various land cover products exhibits substantial variability across LULC categories (Figure 6). Analyzing classification accuracy from 2015 to 2019 reveals consistently high PA values for water, forest, and cropland, whereas grassland, bare ground, and shrubland demonstrate lower and more fluctuating accuracy levels.

Among the evaluated products, the CCI dataset consistently achieves high classification precision for water, with PA values exceeding 0.8, but exhibits challenges in accurately delineating grassland and bare ground. The DW product demonstrates superior performance in the cropland category, achieving near-perfect accuracy ($PA \approx 1$) in 2019, highlighting its strong suitability for capturing specific land cover types in certain years. Similarly, the Esri product performs well in water classification, aligning closely with CCI. The FCS dataset initially exhibited high PA for water in 2015, but its accuracy progressively declined in subsequent years. Meanwhile, the MCD product maintains moderate classification accuracy for water. Overall, FCS and CCI demonstrate relatively stable performance across the GHMA, reflecting the robustness of their classification algorithms. However, the

PA values for grassland, shrubland, and bare ground remain generally low, particularly in the DW and MCD products, with pronounced year-to-year fluctuations.

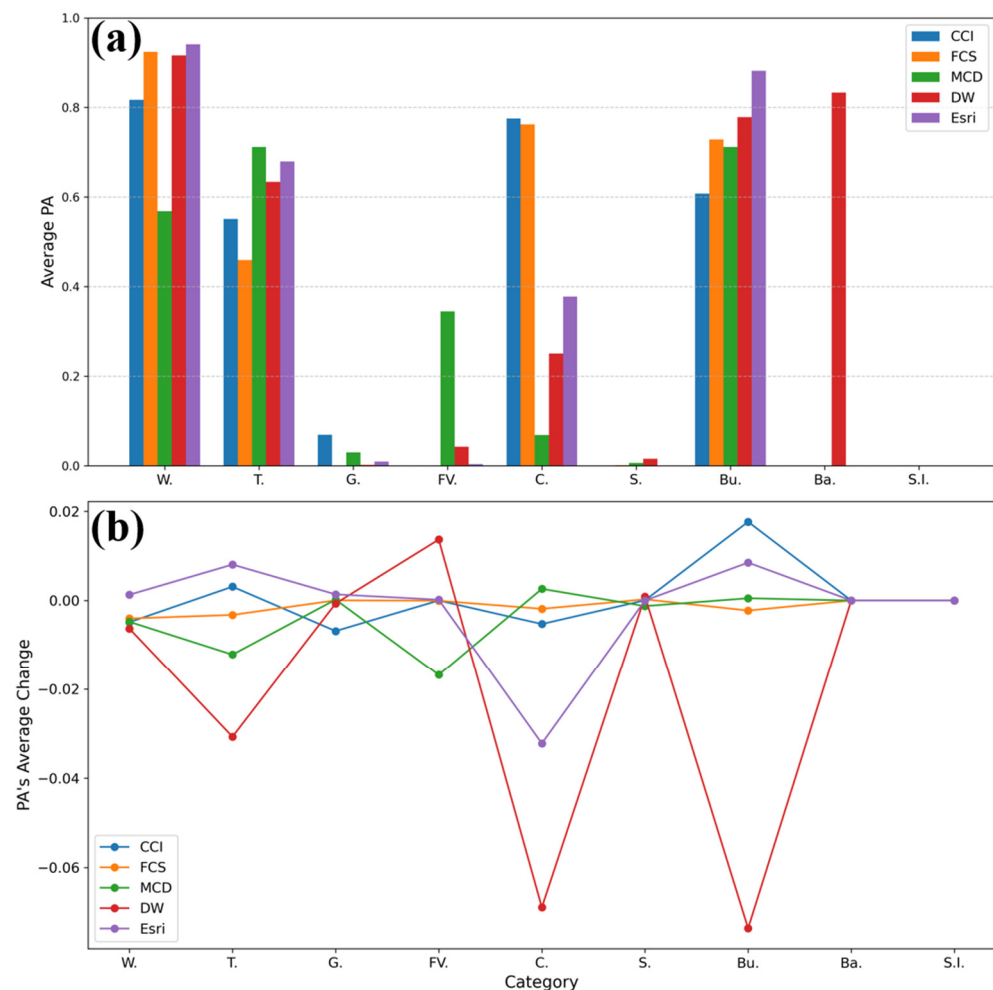


Figure 6. Comparison of PAs by category in the Guangdong–Hong Kong–Macao Greater Bay Area. (a) Average PA; (b) Average Change [‘water’ is ‘W.’, ‘trees’ is ‘T.’, ‘grass’ is ‘G.’, ‘flooded_vegetation’ is ‘FV.’, ‘crops’ is ‘C.’, ‘shrub_and_scrub’ is ‘S.’, ‘built’ is ‘Bu.’, ‘bare’ is ‘Ba.’, ‘snow_and_ice’ is ‘S.I.’].

From a temporal perspective, classification accuracy remained relatively stable between 2015 and 2019, although some categories experienced notable shifts. CCI exhibited minimal variation, whereas FCS showed a gradual decline in water accuracy and a temporary improvement in grassland classification in 2017, followed by a subsequent drop. The DW product recorded a sharp surge in cropland accuracy in 2019 (PA $\approx$ 1), while MCD experienced a substantial decline in cropland accuracy in 2019 (PA $\approx$ 0.2). Despite these variations, classification accuracy for water remained consistently high across all products, underscoring the stability of its spectral characteristics, which facilitate reliable classification.

The Visalia region, situated at the interface of extensive cropland and natural landscapes, is characterized by diverse land cover types and significant anthropogenic influences. Analyzing classification accuracy trends from 2015 to 2019 reveals that most land cover products perform well in cropland and woodland classification, with the CCI dataset demonstrating consistently stable performance within a moderate-to-high accuracy range (Figure 7).

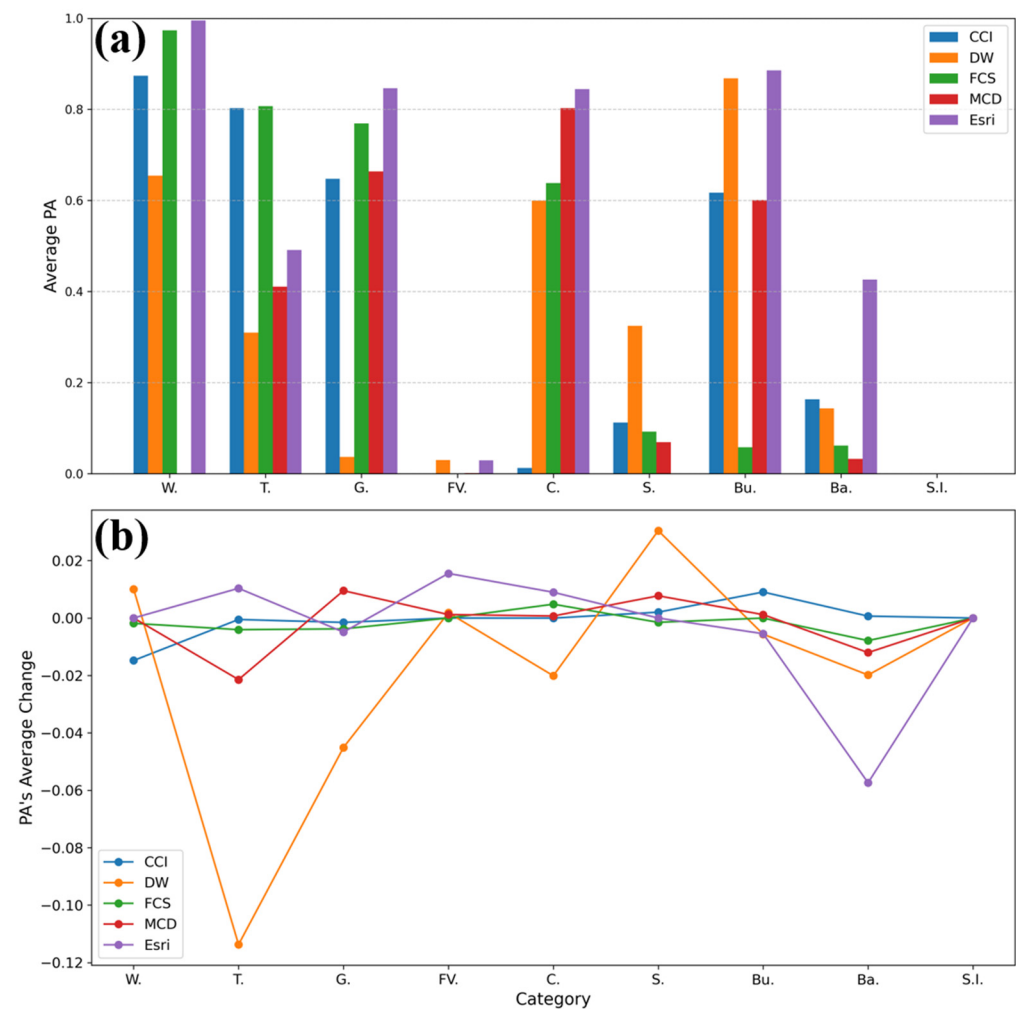


Figure 7. Comparison of PA by category in Visalia, USA. (a) Average PA; (b) Average Change. For explanations of abbreviations, see Figure 6.

Across the study period, the CCI product maintains stable accuracy levels, reinforcing its reliability in land cover classification. The DW product exhibits strong performance, particularly in 2019, when cropland classification accuracy approaches 1, indicating its effectiveness for specific years. Similarly, the Esri product maintains high and stable accuracy across multiple categories, suggesting robust classification capabilities. The FCS dataset performs well, particularly in the tree and grass (likely representing impervious surfaces or built-up areas) categories. Conversely, the MCD product, while achieving moderate accuracy in water classification, demonstrates significant fluctuations in grassland and cropland, indicating variable monitoring capabilities across different land cover types.

Accuracy variations are particularly pronounced among specific land cover categories. Water, trees, and built-up areas exhibit high and stable classification accuracy, suggesting that spectral consistency enhances the ability of these products to distinguish these classes. However, grassland and cropland categories show considerable accuracy fluctuations, likely influenced by complex spectral signatures, seasonal vegetation dynamics, and environmental variability, making them more challenging to classify with high consistency.

The time-series analysis further reveals that classification accuracy in the Visalia region remains relatively stable over time, with CCI and FCS displaying smooth year-to-year variations. However, DW exhibits significant fluctuations, particularly in the tree category, where accuracy is highly unstable, suggesting that DW may not be well suited for dynamic tree cover classification. Conversely, classification accuracy for the built-up

category remains consistently high across products, highlighting their effectiveness in monitoring urban expansion. Nonetheless, classification performance for other land cover categories showed a decline over time, reflecting ongoing challenges in accurately capturing complex landscape dynamics.

The central Norwegian–Swedish border region, characterized by high-latitude woodlands with complex land cover features and strong climatic influences, exhibits generally high classification accuracy in the snow and ice and grassland categories from 2015 to 2019 (Figure 8).

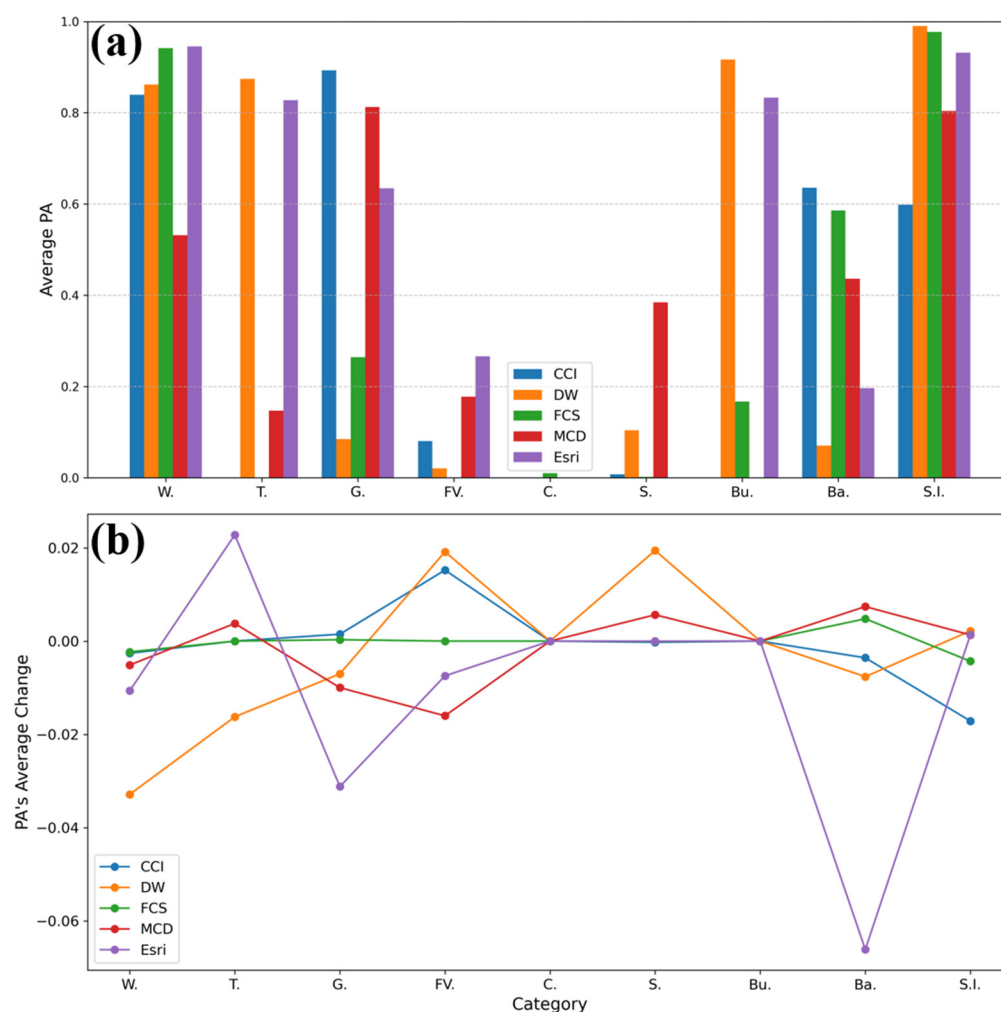


Figure 8. Comparison of PA by category in the central part of the Norwegian–Swedish border. (a) Average PA; (b) Average Change. For explanations of abbreviations, see Figure 6.

Among the evaluated products, CCI demonstrates superior PA in water and tree classifications, while DW performs well across multiple categories, particularly in water. The Esri product maintains consistently high accuracy in the water and tree categories, indicating stable classification performance over multiple years. The FCS product exhibits strong accuracy in water classification but shows fluctuating performance in other categories, suggesting variability in its adaptability to different land cover types. Meanwhile, the MCD product maintains a moderate level of accuracy across categories.

Significant accuracy variations are observed across land cover categories. Most products achieve high and stable classification accuracy in water, reflecting the spectral distinctiveness of this category. However, grassland and cropland classifications display greater

fluctuations, with noticeable accuracy variations across different products, likely due to seasonal vegetation changes and spectral mixing.

From a temporal perspective, classification performance remains relatively stable across products, with year-to-year accuracy changes generally within ± 0.07 . FCS exhibits the most stable performance, while the Esri product shows relatively larger fluctuations over time. Despite these variations, overall classification stability is high in this region, although the Esri product appears less effective in capturing wasteland features.

4.3. Annual Temporal Accuracy

This section evaluates temporal accuracy, as defined in Section 3.2.2, to assess each product's ability to detect and track land cover changes over time.

Figure 9a presents the accuracy of dynamic land cover changes across different time spans for each product in the GHMA. FCS and CCI exhibit the highest accuracy for shorter time spans, indicating their strong capability for high-frequency land cover monitoring. Their ability to capture rapid landscape transitions makes them particularly effective in this region. Conversely, DW consistently exhibits the lowest accuracy across all time spans, suggesting weaker performance in detecting dynamic land cover changes. Esri and MCD demonstrate intermediate accuracy, with notable fluctuations, though they achieve comparable performance to FCS and CCI under certain conditions.

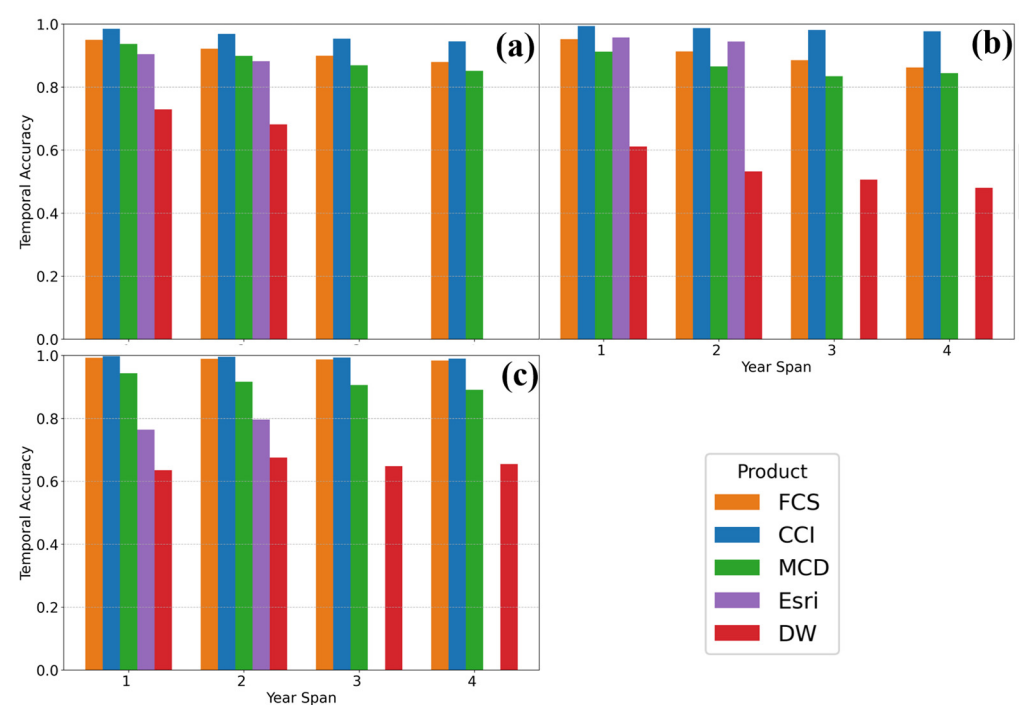


Figure 9. Temporal accuracy of dynamic changes: (a) Guangdong–Hong Kong–Macao Greater Bay Area; (b) Visalia, USA; (c) central Norway–Sweden junction.

In Visalia, USA, FCS maintains high accuracy and stable performance across all time spans, highlighting its adaptability and reliability in monitoring land cover changes. CCI performs well over shorter time spans but experiences a notable accuracy decline as the time span increases, indicating reduced effectiveness for long-term monitoring. Esri outperforms other products for medium-length time spans, exhibiting stable accuracy and robust performance in capturing land cover transitions. MCD remains relatively stable with moderate accuracy, while DW continues to underperform across all time spans, failing to achieve reliable accuracy in any period.

In the central Norwegian–Swedish border region, FCS and CCI products exhibit strong accuracy for shorter time spans but show significant declines in performance over longer periods, making them less suitable for long-term monitoring. Esri, however, maintains a consistent accuracy level across all time spans, demonstrating stability despite not excelling in any particular period. This suggests that Esri is more reliable for long-term monitoring in this region. MCD remains stable with moderate accuracy, exhibiting fewer fluctuations than other products. DW, once again, records the lowest accuracy across all time spans, reinforcing its limited capacity for tracking land cover dynamics in this region.

Across all three study regions, FCS consistently delivers high accuracy in tracking dynamic land cover changes, particularly over shorter time spans, underscoring its versatility and adaptability to diverse monitoring needs. CCI excels in short-term monitoring in the GHMA and Visalia but demonstrates weaker performance in the Norwegian–Swedish border region, suggesting regional variability in its effectiveness. Esri exhibits notable performance variation, excelling in the 2-year time span in Visalia, where it outperforms other products in land cover change detection. MCD maintains a stable, moderate accuracy across all regions, yet lacks exceptional performance in any specific scenario. Finally, DW consistently underperforms in all three regions, with low accuracy across all time spans, highlighting its limitations in capturing land cover change dynamics.

5. Discussion

In Section 4, we presented the results of a comprehensive quantitative analysis of six global time-series land cover data products, assessing their annual classification accuracy, classification performance for different LULC categories, and annual temporal accuracy in three distinct study areas. These analyses revealed the strengths and limitations of each data product under various environmental conditions and LULC types. In this section, these land cover changes will be visually interpreted and discussed together with the results of Section 4 to provide a more intuitive understanding of the actual meaning behind the data and to further explore the applicability of each data product in practical applications.

Considering the rapid urban expansion and diverse LULC types in the GHMA, the FCS, CCI, and DW datasets are suitable choices for monitoring urban expansion and land cover changes. The high resolution (30 m) and refined classification system of FCS effectively identify the details of urban expansion, the long-term data (1992–2024) of CCI are suitable for analyzing long-term trends in urban expansion, and the near-real-time data of DW can monitor the dynamic changes of urban expansion in a timely manner.

In the Visalia region, where land cover types include croplands, grasslands, and urban areas, the MCD and Esri datasets are effective for monitoring urban expansion and land cover changes. MCD provides long-term data (2001–2024) and moderate resolution (500 m), which are suitable for analyzing long-term trends and regional changes in urban expansion, while Esri’s high resolution (10 m) and deep learning technology effectively identify different LULC types, thus accurately monitoring urban expansion.

In the Norway–Sweden border area, dominated by forests and grasslands, the FCS and CCI datasets are effective for monitoring urban expansion and land cover changes. The high resolution (30 m) and refined classification system of FCS can effectively identify forest and grassland types, while the long-term data (1992–2024) of CCI are suitable for analyzing the impact of climate change on forests and grasslands, thereby indirectly monitoring the ecological impact of urban expansion.

5.1. Classification Criteria

A commonality exists among the classification systems of various land cover products, as they all encompass major land cover types. This similarity enables the comparative anal-

ysis of classification performance and temporal variations across products within broadly aligned categories. However, it should be noted that there are also significant differences in category definitions and the degree of subdivision across these systems [30]. For instance, the FCS product differentiates multiple subtypes within the “water” category, including swamp, marsh, flooded flat, and water body, whereas the CCI product consolidates all these subtypes into a single water category. Such discrepancies can lead to classification mismatches, where the same land feature is assigned to different categories across products or where a category encompasses heterogeneous land types. These variations directly impact the reliability of time-series comparisons and trend analyses.

The three-level mapping strategy demonstrates robust performance in categories with pronounced spectral distinctiveness or clearly defined criteria. However, when handling classes with spectral–temporal feature overlap (e.g., grassland/cropland), its effectiveness can be influenced by source data resolution and regional heterogeneity. This limitation is further exemplified by contrasting classification outcomes across regions. In the Norway–Sweden border area, consistently high bare land accuracy reflects the strategy’s strength in distinguishing spectrally distinct features with minimal environmental complexity. Conversely, lower shrubland classification accuracy in the GHMA underscores challenges in resolving mixed vegetation classes within heterogeneous landscapes, where spectral ambiguity and sparse validation data exacerbate mapping uncertainties. Future enhancements should focus on integrating multimodal data and automated rule-based frameworks to improve the strategy’s generalizability and objectivity.

Beyond classification criteria, the mapping process itself affects analysis reliability. The accuracy of category mapping directly influences the reliability of temporal land cover change analyses across multi-source datasets. Misclassification during mapping can introduce two key issues. The first issue is a spatial mismatching problem. For example, misclassifying grass as trees leads to the conclusion of “false forest expansion” [31]. The second one is that of time-series distortion. The process of vegetation gradual change is characterized as abrupt or static in different products due to differences in classification criteria, which interferes with trend analysis. For example, vegetation succession in a region shows gradual change in FCS due to classification refinement, while it shows stability in CGLS due to broad categories, leading to contradictory trends across products. Therefore, when analyzing trends in temporal changes, it is necessary to fully consider the effects of classification mapping and interpret differences and changes between products with caution.

When analyzing temporal land cover trends, it is crucial to account for classification mapping effects and interpret inter-product differences cautiously. A robust mapping framework, combining direct category alignment, spectral feature integration, and expert validation, mitigates misclassification risks and improves the reliability of land cover change assessments.

5.2. Precision Calculation

In land use and land cover time-series change analysis, resampling to a consistent spatial resolution can enhance cross-product monitoring capacity comparisons. However, this process may also introduce accuracy index bias due to spatial distribution distortions and area mismatches [32].

Integrated accuracy evaluation is crucial in land cover studies to ensure the reliability of study results. For classification accuracy, this study employed a randomized 50% window sampling method, averaging results over three iterations, in combination with a fixed 5×5 window strategy. This approach helps mitigate spatial heterogeneity and random noise interference. However, fixed-window sampling may overlook large-scale feature characteristics and mask localized land cover changes, potentially smoothing evaluation

results and reducing sensitivity to small-scale variations [33]. For time-series accuracy, this study utilized a year-by-year comparison approach to assess the consistency of land cover changes across different time scales. This method reveals the long-term performance of classification models through cross-year spanning analysis. However, the use of a 20% random sample size may be insufficient in certain cases, and spatial distribution biases can introduce assessment uncertainty, particularly in spectrally complex regions where classification errors are more likely.

Despite its effectiveness, the current evaluation method has several limitations. The fixed-window approach struggles to balance feature-scale differences. Small windows are susceptible to mixed-pixel effects, while large windows reduce sensitivity to local changes and can obscure finer-scale land cover dynamics. Reliance on a single validation source may lead to spatiotemporal mismatches, affecting the reliability of accuracy assessments. Gradual transitions and mixed-type changes in land cover increase the difficulty of time-series matching, making it challenging to capture nuanced temporal variations.

Therefore, in order to improve the accuracy evaluation, we can improve the following three aspects:

- (1) Dynamic sampling optimization with adaptive window design based on land class spatial heterogeneity and stratified sampling to proportionally allocate samples by type, thereby improving representativeness.
- (2) Multi-source data fusion validation by integrating high-resolution imagery, UAV data, and field surveys to construct a spatiotemporally synchronized reference database, thereby reducing validation bias [34].
- (3) Error-driven parsing and algorithm enhancement by quantifying dominant error sources such as spectral confusion and terrain effects [35,36], and optimizing temporal correlation modeling and change detection mechanisms to enhance the monitoring robustness of gradual processes and mixed land classes.

5.3. Differences in the Products

In the analysis of changes within specific land cover categories, significant variations were observed in the performance of different products across the three study regions: the GHMA, the Visalia region, and the Norway–Sweden border. No single product consistently demonstrated an absolute advantage across all land categories and regions, highlighting the complexity of classification accuracy in diverse environments. The following sections detail the specific differences and their implications.

5.3.1. Environmental and Anthropogenic Influences

In economically developed regions like the GHMA, highly dynamic land cover changes can lead to update lag errors in classification products. The rapid transformation of urban landscapes places higher demands on time-sensitive monitoring [37]. For instance, in the GHMA, products like FCS with high spatial resolution (30 m) are better suited to capture the rapid urban expansion, while DW's near-real-time data help in monitoring the most recent changes. Fluctuations in climate conditions further complicate classification, particularly in grasslands, wetlands, and water bodies, as they influence vegetation phenology rhythms and alter area dynamics, leading to reduced classification consistency [38,39], especially for products like CCI which rely on longer time-series data.

5.3.2. Impact of Data Sources and Sensor Resolution

Differences in data sources, sensor types, and spatial resolution directly impact classification accuracy. Products based on high-resolution satellite imagery, such as FCS (30 m) and DW (10 m), perform well in categories with large, homogeneous areas and distinct spectral

signatures (e.g., water bodies). However, they struggle with complex spectral features found in fragmented landscapes such as grasslands. For example, in the Norway–Sweden border region, the classification of grasslands and forests using FCS can be challenging due to mixed pixels. Low-frequency data updates can result in outdated images, failing to capture sudden land cover changes, which has been a challenge in water body detection for products like CCI that have medium spatial resolution (300 m) and may not capture the most recent changes in dynamic regions [40].

5.3.3. Limitations of Classification Algorithms

Traditional classification methods often lack sensitivity to spectral confusion and environmental variations in complex land cover categories, limiting their generalization ability. While deep learning models like those used in DW have demonstrated robust performance in homogeneous environments, they still exhibit biases in boundary recognition, particularly in fragmented urban landscapes such as bare ground patches [41]. For example, in the Visalia region, products like MCD may struggle with accurately delineating boundaries between cropland and grassland due to similar spectral characteristics.

To enhance classification robustness in complex environments, future research should focus on the following improvements:

- (1) **The Integration of Multi-Source Remote Sensing Data:** Combine optical, radar, and LiDAR data to improve the detection of diverse land cover types under varying environmental conditions. Develop adaptive classification frameworks that incorporate local land cover characteristics. For example, in the GHMA, urban expansion patterns can be better captured by integrating high-resolution optical and radar data to account for frequent cloud cover and rapid changes. Products like FCS and CCI can benefit from such integration to improve their accuracy in complex urban and forested areas.
- (2) **Real-Time Monitoring and Validation:** Integrate Unmanned Aerial Vehicles (UAVs) and ground sensor networks to create an air–sky–ground calibration platform. This system would enable the real-time detection of surface changes, helping to reduce update lag errors in time-series monitoring, especially in dynamic regions like the GHMA where urban expansion is rapid. DW’s near-real-time data can be further enhanced through such a platform to provide more accurate and timely updates.
- (3) **Advanced Algorithm Development:** Improve deep learning-based classification models by enhancing their ability to differentiate between spectrally similar land cover types and accurately delineate boundaries in heterogeneous landscapes. Develop hybrid models that incorporate physical-based and data-driven approaches to improve classification stability and accuracy, particularly in regions with complex land cover such as the Norway–Sweden border. For instance, combining the strengths of FCS’s high resolution with advanced algorithms could lead to the better classification of forest and grassland boundaries in this region.

6. Conclusions

This study undertakes a comprehensive analysis of six recent time-series land cover datasets, incorporating the critical metric of time-series accuracy to evaluate their performance across three representative regions. The evaluation framework integrates both classification accuracy and time-series accuracy assessments, enabling a robust and systematic comparison of multi-source, long-time-series land cover data.

The comparison results indicate that while the datasets exhibit spatial consistency, significant discrepancies exist in land cover classification. In terms of accuracy performance, each dataset demonstrates varying levels of accuracy depending on the region and land cover type, with no single dataset outperforming others in all case study areas. FCS is good

for short-term forest and urban expansion monitoring due to its high versatility and stability in short-term dynamic changes, but needs improvement in long-term monitoring and complex land cover identification. CCI is characterized by long-term continuity, medium resolution, and stable classification accuracy, and is suitable for scenarios requiring long-term land cover change analysis (e.g., climate modeling, ecological trend studies), but in complex terrain or fragmented land cover areas, it needs to be combined with high-resolution data to improve accuracy. Esri is recommended for high-resolution analysis in specific regions like urban planning and agricultural areas due to its excellent performance in certain time spans and regions, but has limited adaptability elsewhere. MCD maintains moderate accuracy across regions, suitable for medium-resolution regional land cover analysis. DW, while generally underperforming in dynamic land cover change monitoring, has some value in the preliminary monitoring of rapidly changing areas due to its near-real-time data. For dataset selection, with reference to the regional context, the following recommendations are made:

In the GHMA, FCS, CCI, and DW are recommended. FCS's high resolution (30 m) and detailed classification system effectively capture urban expansion details. CCI's long-term data are suitable for analyzing long-term trends. DW's near-real-time data can monitor urban expansion dynamics in a timely manner.

In the Visalia region of the United States, MCD and Esri are suggested. MCD's long-term data and medium resolution are good for analyzing long-term trends and regional changes. Esri's high resolution and deep learning technology effectively identify different land cover types.

In the Norway–Sweden border region, FCS and CCI are recommended. FCS's high resolution and detailed classification system effectively identify forest and grassland types. CCI's long-term data and medium resolution are suitable for analyzing long-term trends and the impact of climate change.

While the recommendations above are derived from rigorous analysis across three diverse transitional regions, we recognize that their applicability to other global contexts (e.g., arid zones, tropical rainforests, or small island states) may vary. These case studies provide a foundation for context-driven dataset selection but do not fully capture the complexity of all landscapes. We therefore encourage future research to expand validation efforts to underrepresented regions, such as sub-Saharan Africa, the Amazon Basin, or the Tibetan Plateau, to further refine and generalize these findings. Such work would strengthen the global applicability of land cover dataset selection strategies and address remaining gaps in heterogeneous environments.

While contemporary land cover products provide valuable insights for long-term monitoring, they face three critical constraints: semantic inconsistencies in class definitions, identification errors across complex land surfaces, and limited sensitivity to transitional dynamics. To address these challenges, next-generation data development requires standardized hierarchical taxonomies with ecoregion-adapted ontologies to ensure multi-scale representational fidelity. Concurrent technological integration should combine multi-sensor synergies (optical/SAR/LiDAR), implement physics-constrained super-resolution architectures for temporal coherence, and embed process-aware mechanisms through coupled land–atmosphere modeling. This integrated approach targets enhanced spatiotemporal continuity and classification logic integrity across heterogeneous landscapes, particularly improving transitional zone characterization and change detection reliability in peri-urban/ecotone regions.

Author Contributions: Conceptualization, P.L., C.W. and C.Z.; methodology, P.L., Y.W., L.T., C.W. and C.Z.; validation, C.W., Y.W., M.L. and S.X.; formal analysis, P.L., C.W. and Y.W.; investigation, P.L., C.W. and C.Z.; writing—original draft preparation, P.L. and Y.W.; writing—review and editing, P.L.,

Y.W. and C.Z.; visualization, Y.W., M.L. and S.X.; supervision, C.W. and C.Z.; project administration, C.W. and C.Z.; funding acquisition, C.W. and C.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China [42374018; 42304012], the Guangdong Basic and Applied Basic Research Foundation [2022A1515110730; 2025B1515020092], the Shenzhen Science and Technology Program [KCXFZ20240903093000002; JCYJ20220531101409021].

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

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

References

1. Yang, Y.; Xiao, P.; Feng, X.; Li, H. Accuracy Assessment of Seven Global Land Cover Datasets over China. *ISPRS J. Photogramm. Remote Sens.* **2017**, *125*, 156–173. [\[CrossRef\]](#)
2. Song, X.-P.; Huang, C.; Feng, M.; Sexton, J.O.; Channan, S.; Townshend, J.R. Integrating Global Land Cover Products for Improved Forest Cover Characterization: An Application in North America. *Int. J. Digit. Earth* **2014**, *7*, 709–724. [\[CrossRef\]](#)
3. Xiao, C.; Li, P.; Feng, Z. Agricultural Expansion and Forest Retreat in Mainland Southeast Asia since the Late 1980s. *Land Degrad. Dev.* **2023**, *34*, 5606–5621. [\[CrossRef\]](#)
4. Chinmayi, H.K.; Flynn, K.C.; Ashworth, A.J. Advancements in Remote Sensing Techniques for Earthquake Engineering: A Review. *Earthq. Res. Adv.* **2024**, *4*, 100352. [\[CrossRef\]](#)
5. Thackway, R.; Lymburner, L.; Guerschman, J.P. Dynamic Land Cover Information: Bridging the Gap between Remote Sensing and Natural Resource Management. *Ecol. Soc.* **2013**, *18*, 2. [\[CrossRef\]](#)
6. Yang, G.; Fang, S.; Gong, W.; Zhao, Y.; Ge, M. Evaluating the Reliability of Time Series Land Cover Maps by Exploiting the Hidden Markov Model. *Stoch. Environ. Res. Risk Assess.* **2021**, *35*, 881–892. [\[CrossRef\]](#)
7. Tsutsumida, N.; Comber, A.J. Measures of Spatio-Temporal Accuracy for Time Series Land Cover Data. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *41*, 46–55. [\[CrossRef\]](#)
8. Herold, M.; Mayaux, P.; Woodcock, C.E.; Baccini, A.; Schmullius, C. Some Challenges in Global Land Cover Mapping: An Assessment of Agreement and Accuracy in Existing 1 Km Datasets. *Remote Sens. Environ.* **2008**, *112*, 2538–2556. [\[CrossRef\]](#)
9. Tsendbazar, N.E.; De Bruin, S.; Herold, M. Assessing Global Land Cover Reference Datasets for Different User Communities. *ISPRS J. Photogramm. Remote Sens.* **2015**, *103*, 93–114. [\[CrossRef\]](#)
10. Moody, A.; Woodcock, C.E. Scale-Dependent Errors in the Estimation of Land-Cover Proportions. Implications for Global Land-Cover Datasets. *Photogramm. Eng. Remote Sens.* **1994**, *60*, 585–594.
11. Tudesque, L.; Tisseuil, C.; Lek, S. Scale-Dependent Effects of Land Cover on Water Physico-Chemistry and Diatom-Based Metrics in a Major River System, the Adour-Garonne Basin (South Western France). *Sci. Total Environ.* **2014**, *466–467*, 47–55. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Waśniewski, A.; Hościło, A.; Aune-Lundberg, L. The Impact of Selection of Reference Samples and DEM on the Accuracy of Land Cover Classification Based on Sentinel-2 Data. *Remote Sens. Appl. Soc. Environ.* **2023**, *32*, 101035. [\[CrossRef\]](#)
13. Abercrombie, S.P.; Friedl, M.A. Improving the Consistency of Multitemporal Land Cover Maps Using a Hidden Markov Model. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 703–713. [\[CrossRef\]](#)
14. Yang, H.; Li, S.; Chen, J.; Zhang, X.; Xu, S. The Standardization and Harmonization of Land Cover Classification Systems towards Harmonized Datasets: A Review. *ISPRS Int. J.-Geo-Inf.* **2017**, *6*, 154. [\[CrossRef\]](#)
15. Liu, H.; Gong, P.; Wang, J.; Clinton, N.; Bai, Y.; Liang, S. Annual Dynamics of Global Land Cover and Its Long-Term Changes from 1982 to 2015. *Earth Syst. Sci. Data* **2020**, *12*, 1217–1243. [\[CrossRef\]](#)
16. Fritz, S.; See, L.; McCallum, I.; Schill, C.; Obersteiner, M.; Van Der Velde, M.; Boettcher, H.; Havlík, P.; Achard, F. Highlighting Continued Uncertainty in Global Land Cover Maps for the User Community. *Environ. Res. Lett.* **2011**, *6*, 044005. [\[CrossRef\]](#)
17. Zhang, X.; Zhao, T.; Xu, H.; Liu, W.; Wang, J.; Chen, X.; Liu, L. GLC_FCS30D: The First Global 30 m Land-Cover Dynamics Monitoring Product with a Fine Classification System for the Period from 1985 to 2022 Generated Using Dense-Time-Series Landsat Imagery and the Continuous Change-Detection Method. *Earth Syst. Sci. Data* **2024**, *16*, 1353–1381. [\[CrossRef\]](#)
18. Karra, K.; Kontgis, C.; Statman-Weil, Z.; Mazzariello, J.C.; Mathis, M.; Brumby, S.P. Global Land Use/Land Cover with Sentinel 2 and Deep Learning. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 4704–4707.
19. Sulla-Menashe, D.; Gray, J.M.; Abercrombie, S.P.; Friedl, M.A. Hierarchical Mapping of Annual Global Land Cover 2001 to Present: The MODIS Collection 6 Land Cover Product. *Remote Sens. Environ.* **2019**, *222*, 183–194. [\[CrossRef\]](#)

20. Liu, X.; Yu, L.; Si, Y.; Zhang, C.; Lu, H.; Yu, C.; Gong, P. Identifying Patterns and Hotspots of Global Land Cover Transitions Using the ESA CCI Land Cover Dataset. *Remote Sens. Lett.* **2018**, *9*, 972–981. [\[CrossRef\]](#)
21. Mousivand, A.; Arsanjani, J.J. Insights on the Historical and Emerging Global Land Cover Changes: The Case of ESA-CCI-LC Datasets. *Appl. Geogr.* **2019**, *106*, 82–92. [\[CrossRef\]](#)
22. Brown, C.F.; Brumby, S.P.; Guzder-Williams, B.; Birch, T.; Hyde, S.B.; Mazzariello, J.; Czerwinski, W.; Pasquarella, V.J.; Haertel, R.; Ilyushchenko, S.; et al. Dynamic World, Near Real-Time Global 10 m Land Use Land Cover Mapping. *Sci. Data* **2022**, *9*, 251. [\[CrossRef\]](#)
23. Buchhorn, M.; Lesiv, M.; Tsendbazar, N.-E.; Herold, M.; Bertels, L.; Smets, B. Copernicus Global Land Cover Layers—Collection 2. *Remote Sens.* **2020**, *12*, 1044. [\[CrossRef\]](#)
24. Bestelmeyer, B.T.; Wiens, J.A. Local and Regional-scale Responses of Ant Diversity to a Semiarid Biome Transition. *Ecography* **2001**, *24*, 381–392. [\[CrossRef\]](#)
25. Li, W.; Fan, J.; Li, Z.; Wang, C.; Zhang, X.; Duan, J. Improved Adaptive Neuro-fuzzy Inference System with Bacterial Foraging Optimization Algorithm for Suspended Sediment Concentration Estimation. *J. Interll. Fuzzy Syst.* **2024**, *46*, 3945–3961. [\[CrossRef\]](#)
26. Wang, C.; Zhu, C.; Wang, X.; Tu, W.; Li, Q. A New Object-Oriented SAR Interferometry Framework for Monitoring Urban Deformation. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5227011. [\[CrossRef\]](#)
27. Xie, S.; Liu, L.; Zhang, X.; Chen, X. Annual Land-Cover Mapping Based on Multi-Temporal Cloud-Contaminated Landsat Images. *Int. J. Remote Sens.* **2019**, *40*, 3855–3877. [\[CrossRef\]](#)
28. Wang, X.; Cao, J.; Liu, J.; Li, X.; Wang, L.; Zuo, F.; Bai, M. Improving the Interpretability and Reliability of Regional Land Cover Classification by U-Net Using Remote Sensing Data. *Chin. Geogr. Sci.* **2022**, *32*, 979–994. [\[CrossRef\]](#)
29. Higgins, S.I.; Conradi, T.; Muhoko, E. Shifts in Vegetation Activity of Terrestrial Ecosystems Attributable to Climate Trends. *Nat. Geosci.* **2023**, *16*, 147–153. [\[CrossRef\]](#)
30. Zheng, Q.-H.; Chen, W.; Li, S.-L.; Yu, L.; Zhang, X.; Liu, L.-F.; Singh, R.P.; Liu, C.-Q. Accuracy Comparison and Driving Factor Analysis of LULC Changes Using Multi-Source Time-Series Remote Sensing Data in a Coastal Area. *Ecol. Inform.* **2021**, *66*, 101457. [\[CrossRef\]](#)
31. Alganci, U. Dynamic Land Cover Mapping of Urbanized Cities with Landsat 8 Multi-Temporal Images: Comparative Evaluation of Classification Algorithms and Dimension Reduction Methods. *ISPRS Int. J.-Geo-Inf.* **2019**, *8*, 139. [\[CrossRef\]](#)
32. Kim, D.-H.; Johnson, J.M.; Clarke, K.C.; McMillan, H.K. Untangling the Impacts of Land Cover Representation and Resampling in Distributed Hydrological Model Predictions. *Environ. Model. Softw.* **2024**, *172*, 105893. [\[CrossRef\]](#)
33. Wang, H.; Yan, H.; Hu, Y.; Xi, Y.; Yang, Y. Consistency and Accuracy of Four High-Resolution LULC Datasets—Indochina Peninsula Case Study. *Land* **2022**, *11*, 758. [\[CrossRef\]](#)
34. Qu, L.; Chen, Z.; Li, M.; Zhi, J.; Wang, H. Accuracy Improvements to Pixel-Based and Object-Based LULC Classification with Auxiliary Datasets from Google Earth Engine. *Remote Sens.* **2021**, *13*, 453. [\[CrossRef\]](#)
35. Parracciani, C.; Gigante, D.; Mutanga, O.; Bonafoni, S.; Vizzari, M. Land Cover Changes in Grassland Landscapes: Combining Enhanced Landsat Data Composition, LandTrendr, and Machine Learning Classification in Google Earth Engine with MLP-ANN Scenario Forecasting. *GIScience Remote Sens.* **2024**, *61*, 2302221. [\[CrossRef\]](#)
36. Huang, X.; Ibrahim, M.M.; Luo, Y.; Jiang, L.; Chen, J.; Hou, E. Land Use Change Alters Soil Organic Carbon: Constrained Global Patterns and Predictors. *Earth's Future* **2024**, *12*, e2023EF004254. [\[CrossRef\]](#)
37. Fu, P.; Weng, Q. A Time Series Analysis of Urbanization Induced Land Use and Land Cover Change and Its Impact on Land Surface Temperature with Landsat Imagery. *Remote Sens. Environ.* **2016**, *175*, 205–214. [\[CrossRef\]](#)
38. Kaiser, E.A.; Rolim, S.B.A.; Grondona, A.E.B.; Hackmann, C.L.; De Marsillac Linn, R.; Käfer, P.S.; Da Rocha, N.S.; Diaz, L.R. Spatiotemporal Influences of LULC Changes on Land Surface Temperature in Rapid Urbanization Area by Using Landsat-TM and TIRS Images. *Atmosphere* **2022**, *13*, 460. [\[CrossRef\]](#)
39. Zheng, H.; Chen, Y.; Pan, W.; Cai, Y.; Chen, Z. Impact of Land Use/Land Cover Changes on the Thermal Environment in Urbanization: A Case Study of the Natural Wetlands Distribution Area in Minjiang River Estuary, China. *Pol. J. Environ. Stud.* **2019**, *28*, 3025–3041. [\[CrossRef\]](#)
40. Sagan, V.; Peterson, K.T.; Maimaitijiang, M.; Sidike, P.; Sloan, J.; Greeling, B.A.; Maalouf, S.; Adams, C. Monitoring Inland Water Quality Using Remote Sensing: Potential and Limitations of Spectral Indices, Bio-Optical Simulations, Machine Learning, and Cloud Computing. *Earth-Sci. Rev.* **2020**, *205*, 103187. [\[CrossRef\]](#)
41. Arvor, D.; Durieux, L.; Andrés, S.; Laporte, M.-A. Advances in Geographic Object-Based Image Analysis with Ontologies: A Review of Main Contributions and Limitations from a Remote Sensing Perspective. *ISPRS J. Photogramm. Remote Sens.* **2013**, *82*, 125–137. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.