remote sensing

MDPI

Article

Multi-Decadal Remote Sensing of Land Cover and Surface Water Dynamics in a Data-Limited Monsoon Lake Watershed in Myanmar

La Minn Ko Ko [1], Win Sithu Maung [2,*] and Phyo Thandar Hlaing [3]

[1] Independent Researcher, Tsukuba 305-0035, Ibaraki, Japan; laminnkoko.forestry@gmail.com
[2] Japan International Research Center for Agricultural Sciences (JIRCAS), 1-1 Ohwashi, Tsukuba 305-8686, Ibaraki, Japan
[3] Disaster Prevention Research Institute, Kyoto University, Gokasho, Uji 611-0011, Kyoto, Japan; phyothandar.hlaing.3m@kyoto-u.ac.jp
* Correspondence: wins0323@jircas.go.jp

Highlights

What are the main findings?

- Under spatially grouped internal testing, SVM achieved the highest 26-year mean overall accuracy (71.49%), kappa (0.620) and macro-F1 (71.79%) among the three classifiers evaluated.
- Reference-based estimates supported agricultural expansion and forest loss during 2000–2025, while watershed-wide surface water change remained uncertain. Annual trend significance for Agriculture and Built-up classes depended on the assumed classification error corrections. Within the Landsat 8-only period (2013–2025), mapped Agriculture showed no monotonic trend.

What are the implications of the main findings?

- Interpreting watershed change requires reference-based net change estimates and their sampling uncertainty alongside annual mapped trajectories.
- Using a fixed common domain makes annual mapped area and water transition accounting directly comparable.

Abstract

Long-term monitoring of land use/land cover (LULC) and surface water is challenging in monsoon lake watersheds with limited reference observations. This study constructed annual LULC in the Inle lake watershed, Myanmar, for 2000–2025 using Landsat annual median composites, six reflectance bands and six spectral indices. Random Forest, Support Vector Machine (SVM) and XGBoost were evaluated using spatially grouped internal testing of purposively selected reference samples. SVM achieved the highest 26-year mean internal overall accuracy (71.49%), kappa (0.620) and macro-F1 (71.79%). A separate stratified probability sample of 600 fixed pixels was assessed for 2000, 2010, 2020 and 2025 to estimate map accuracy, class areas and paired net changes within a common 5438.20 km^2 domain. Area-weighted overall accuracy ranged from 58.7% to 71.4%, while Built-up user's accuracy remained low at 12.4–22.9%. Reference-based 2000–2025 changes were Agriculture +450.1 km^2 (95% sampling interval: +203.4 to +696.8), Forest −631.5 km^2 (−857.3 to −405.7) and Water +42.3 km^2 (−23.5 to +108.0). The Water interval included zero, and the reference assessment did not establish a 2020–2025 recovery. Annual trend significance for Agriculture and Built-up changed under assumed classification error correction scenarios. Within the Landsat 8-only period (2013–2025), mapped Forest still declined, whereas

Check for updates

Academic Editors: Jingzhe Wang, Xiangyu Ge, Zipeng Zhang, Ivan Lizaga, Yinghui Zhang and Yangyi Wu

Received: 24 September 2026
Revised: 6 November 2026
Accepted: 6 November 2026
Published: 8 November 2026

Copyright: © 2026 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.

mapped Agriculture showed no monotonic trend. Within the fixed open water outline of Inle lake, mapped water declined slowly but consistently (-0.07 km^2 yr^{-1}; $p < 0.001$), the Moebye reservoir showed no trend, and most of the mapped decline occurred in the surrounding sanctuary wetlands. Most of the reference-based agricultural increase and forest loss occurred between 2000 and 2010. The results support agricultural expansion and forest loss but leave watershed-wide surface water change uncertain. Distinguishing raw mapped trajectories from reference-based estimates and sampling uncertainty is essential for interpreting long-term watershed change.

Keywords: surface water; LULC; Landsat satellite imagery; Support Vector Machine; Random Forest; XGBoost; Inle lake watershed

1. Introduction

Lakes are key components of watershed systems. They store freshwater, regulate hydrological processes, support biodiversity, and sustain human livelihoods. In many developing regions, they also provide essential ecosystem services related to agriculture, fisheries, transportation, tourism, and cultural heritage [1,2]. Their multiple roles make lakes the focal points of both ecological and socio-economic activities. However, lake systems are increasingly affected by climate variability and human-induced change in land use. Alterations in land cover, water use, and climate can modify runoff generation, sediment delivery, evaporation, water quality, and the spatial extent of surface water [3,4]. Hence, monitoring these interconnected changes is indispensable for understanding lake vulnerability and supporting sustainable watershed management.

The Inle lake watershed in southern Shan State in eastern Myanmar is an ecologically and socio-economically important freshwater system. Inle lake and its associated wetlands support fisheries, floating and terrestrial agriculture, transportation, tourism, and local livelihoods. Likewise, the watershed has experienced long-standing concern over agricultural expansion, forest degradation, sedimentation, water quality deterioration, and modification of lake margin environments [1,5–7]. Sidle et al. [7] documented the historical loss of open-water area in Inle lake and emphasized the role of in-lake and near-lake agricultural practices, whereas subsequent research work has highlighted broader change in land cover, degradation of ecosystem service, and complex interactions between humans and water across the catchment [1,5,6].

Satellite remote sensing is an effective technology to monitor changes in land use/ land cover and hydrology where continuous hydrometric and field observations are limited. The Landsat archive is especially invaluable since it offers a multi-decadal record at 30 m spatial resolution with broadly consistent spectral information [8]. Cloud computing environments such as Google Earth Engine (GEE) allow for large Landsat archives to be filtered, composited, processed, and reproduced [9]. Pixel-based compositing is particularly useful in cloud-prone regions, for it can reduce the influence of missing observations and individual acquisition dates [10]. At the global scale, Landsat has been used to quantify the occurrence, persistence, gain, and loss of surface water over multiple decades [11,12]. Although these technological advances in image processing provide a methodological foundation for watershed-scale analysis, local studies are still necessary because shallow water, floating vegetation, wetlands, cultivated margins, and mixed pixels can behave differently from open-water systems.

Machine learning classification offers a flexible method to integrate spectral bands and diagnostic indices when class boundaries are nonlinear. Random Forest (RF), Support Vector Machine (SVM), and gradient-boosting methods are widely used in remote

sensing since they can handle multidimensional predictor spaces and complex spectral relationships [13–15]. However, the apparent performance of a classifier can be inflated when training and validation samples are spatially dependent. Spatially structured validation is therefore preferable when reference observations are geographically clustered, or neighboring pixels share similar spectral and environmental properties [16]. In the case of multi-year studies, it is equally important not to select a classifier from a single reference year and simply assume that the selected classifier is transferred without changing over the decades as sensor differences and evolving landscape conditions can alter class separability.

This study complements earlier lake-focused and selected year assessments [5–7] by providing an annual watershed-wide land cover record for 2000–2025 and explicitly separating mapped trajectories from reference-supported change estimates. Its main methodological contribution is to pair an annual, spatially grouped comparison of three classifiers, repeated in every year of a 26-year record, with design-based estimates of area and change from a stratified probability sample, so that mapped trends can be checked against estimates that are not biased by map error. Predictor ablation, seasonal compositing comparisons and trend sensitivity analyses examine how methodological choices affect the interpretation of the resulting record. This combination allows for agricultural expansion, forest loss and apparent surface water changes to be assessed alongside classification error, sampling uncertainty and observational limitations. The contribution is an integrated assessment of the evidence supporting watershed change, rather than a new classification algorithm or independent validation of every annual map. Reference-based uncertainty estimates remain conditional on the validity of the interpreted labels.

This study assesses annual land cover and watershed-wide surface water patterns in the Inle lake watershed during 2000–2025. Its objectives are to (1) compare RF, SVM and XGBoost using spatially grouped internal evaluation and select the production classifier using Development-only cross-validation; (2) quantify raw mapped annual class areas and surface water persistence, gain and loss within a fixed common valid domain; (3) estimate production map accuracy, reference-based class areas and paired net changes for 2000, 2010, 2020 and 2025 using a stratified probability reference panel, with sampling uncertainty conditional on reference label validity; and (4) examine annual trend sensitivity to lag-1 dependence and assumed classification error correction patterns. The annual maps and reference-based estimates are interpreted jointly to describe watershed change and identify priorities for further validation. Attribution to specific climatic, agricultural or reservoir management drivers remains outside the scope of this analysis.

2. Materials and Methods

2.1. Study Site

The Inle lake watershed is situated in southern Shan State in eastern Myanmar within the upper Thanlwin/Salween River basin (Figure 1). Inle lake is a shallow freshwater lake located at approximately 875–890 m above sea level. Its tributaries comprise several small streams draining the surrounding uplands, while outflow occurs through the Balu Chaung. Due to its shallow depth and low-relief margins, the lake is sensitive to rainfall variability, runoff, evaporation, sedimentation, and land use pressure [7,17]. The delineated study boundary (Figure 1b) also includes Moebye Dam and its reservoir, a hydropower impoundment on the Balu Chaung immediately downstream of the lake's outlet. Reservoir operations are a potential influence on the aggregate Water class, but their contribution was not quantified in this study. This is discussed further in Section 4.4.

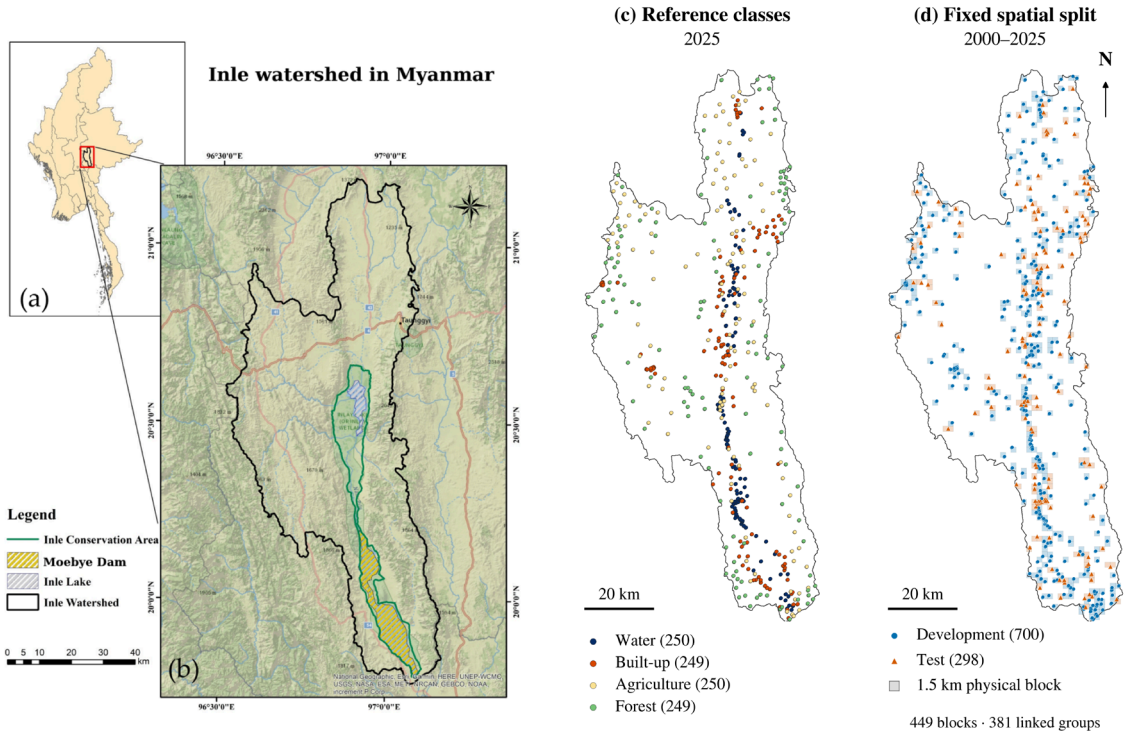

Figure 1. Location and reference sample design of the Inle lake watershed, Myanmar: (**a**) location within Myanmar; (**b**) watershed boundary, Inle lake, Moebye reservoir and the conservation area; (**c**) the 998 quality-controlled purposive reference locations with their 2025 class labels (Water, 250; Built-up, 249; Agriculture, 250; Forest, 249); and (**d**) the fixed Development (700 locations) and Test (298 locations) assignment, with the 449 occupied 1.5 km square physical blocks. The locations and outer split were retained across all 26 years, 2000–2025, while class labels were checked by year and updated when interpreted land cover changed. Panels (**c**,**d**) use WGS84/UTM Zone 47N (EPSG:32647), identical extents and display scales, and the rasterized watershed outline; the north arrow applies to both panels. Occupied blocks are shown in full and may extend beyond the outline. Panels (**c**,**d**) depict the purposive classifier reference pool; the separate 600-pixel probability reference panel is described in Section 2.6.

The Inle lake watershed has a monsoonal climate consisting of a wet season from May to October and a dry season from November to April. Seasonal rainfall mainly controls surface runoff, floodplain inundation, groundwater recharge, sediment transport, and lake expansion or contraction. Temperature and evapotranspiration also influence the surface water balance, particularly during dry season when reduced rainfall and higher evaporative demand may contribute to a decline in surface water.

Land use in the watershed reflects human–environment interaction in the long term. Agriculture in lowland and moderately sloping areas is the predominant land use activity such as upland farming and floating agriculture around the lake. Forest cover is mainly distributed in the surrounding uplands. However, deforestation, agricultural expansion, settlement growth, tourism development, and infrastructure expansion have increased pressure on land and water resources. They also have been linked to sedimentation, altered runoff processes, ecosystem service degradation, and declining environmental quality [1,6,17,18].

This setting motivates an assessment of annual land cover and watershed-wide mapped surface water patterns, with explicit consideration of classification and reference uncertainty. The analysis describes these patterns without attributing them to monsoon variability, land use activities or reservoir operations.

2.2. Landsat Data and Preprocessing

Annual Landsat surface reflectance imagery was prepared in GEE using Landsat Collection 2 Level 2 products. The Landsat 5 TM collection was used for 2000–2001 and 2004–2011. For 2002, 2003 and 2012, Landsat 5 TM and Landsat 7 ETM+ collections were queried, and their available preprocessed observations were merged before compositing. The Landsat 8 OLI collection was used for 2013–2025. Landsat 5 TM was planned for 2000–2012 and Landsat 8 OLI for 2013–2025. No Landsat 5 TM images were available for the study area in 2012, because Landsat 5 TM imaging ended in November 2011, so Landsat 7 ETM+ was used for that year. In 2002 and 2003, Landsat 5 TM observations alone did not provide sufficient coverage, so Landsat 7 ETM+ observations were added. These designations describe the eligible sensor collections; actual contributions depended on the observations remaining after spatial, temporal and quality filtering and did not necessarily include both legacy sensors in every designated year. For each year, scenes intersecting the watershed were filtered to a scene-level, cloud cover threshold of less than 30%. Pixel quality masking used QA_PIXEL bits 0, 1, 3, 4 and 5 for Landsat 5 TM and Landsat 7 ETM+, excluding fill, dilated cloud, cloud, cloud shadow and snow, respectively. For Landsat 8 OLI, bits 0–5 were masked, additionally excluding cirrus. Optical saturation was screened using QA_RADSAT bits 0–4 and 6 for TM/ETM+ and bits 1–6 for OLI. Legacy images were additionally screened using QA_RADSAT bit 9 for dropped pixels. The six optical bands were converted to surface reflectance using DN × 0.0000275 − 0.2, and annual median composites were calculated from valid January–December observations. This annual compositing strategy was chosen to maximize spatial completeness and temporal consistency across the multi-decadal archive rather than represent a single hydrological season. The resulting annual Water map is therefore neither a maximum nor a minimum inundation extent and is not directly comparable with seasonal lake surface statistics reported elsewhere [8–10]. A calendar year (January–December) window was used rather than a hydrological water year definition. In view of the watershed's wet season (May–October, Section 2.1) falling entirely within a single calendar year, this choice does not split a monsoon season across two annual composites. The complete workflow is summarized in Figure 2.

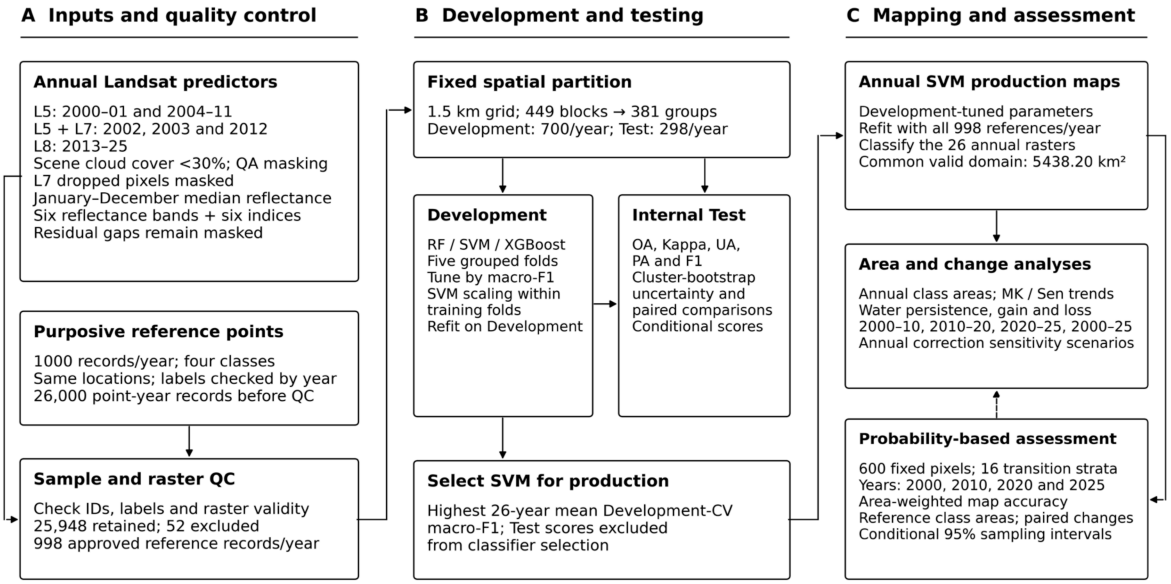

Figure 2. Workflow for Landsat-based LULC mapping and surface water analysis in the Inle lake Watershed (2000–25).

Landsat 7 ETM+ dropped pixels flagged by bit 9 of QA_RADSAT were masked before compositing, together with pixels excluded by the cloud, shadow, snow, fill and optical saturation masks. For 2002, 2003 and 2012, the preprocessed Landsat 5 and Landsat 7 collections were merged, and annual median reflectance was calculated from the available valid observations. Residual gaps without valid observations remained masked. Raster quality control excluded samples with missing predictor values and mapped area and change comparisons were restricted to the fixed common valid domain. A scene inventory was compiled from the Landsat archive to document acquisition availability by year, sensor and season (Table A1). In this inventory, the qualifying observations for 2002 and 2012 were exclusively from Landsat 7 ETM+, although the queried collection pools included both Landsat 5 and Landsat 7. Dry season scene counts exceeded wet season counts in all 26 years. These counts describe scene availability; pixel-specific observation availability additionally depends on quality masking and scene coverage. The complete inventory, including acquisition dates, sensor identifiers, Landsat product identifiers, path/row information and scene-level cloud cover, is provided in Supplementary Data S1.

All annual predictor stacks were clipped to the same watershed domain and exported as GeoTIFF files for analysis in Python 3.14.7 (Google Colab). Spectral bands and indices that were conceptually equivalent were maintained across the Landsat sensors so that the annual models used a consistent definition of predictor. No empirical cross-sensor bandpass transformation or overlap-based radiometric calibration was applied. Consistent band definitions and scaling do not establish radiometric equivalence among sensors.

Seasonal compositing sensitivity was assessed for 2000, 2012 and 2025 using annual (January–December), dry season (January–April plus November–December of the same calendar year) and wet season (May–October) median composites. Sensor selection, quality masking and predictor calculation procedures followed the annual workflow. For each year, the production SVM fitted to the 998 approved annual reference samples, using parameters selected within the Development data, was held fixed and applied to the seasonal predictors without seasonal retuning.

All comparisons used one fixed domain containing valid twelve-predictor values for every year composite combination within the full common domain. A consistency check found 15 pixels in 2012 (0.01257 km^2) where the annual reflectance used for the seasonal comparison differed from that of the production stack; these locations were excluded from all nine comparisons. At all other locations, the annual reflectance was identical to the production stack. Temporal change was compared between 2000 and 2025, while 2012 provided an intermediate compositing and coverage diagnostic. This descriptive experiment assessed sensitivity to compositing season; seasonal classification accuracy was not independently evaluated.

2.3. Predictor Variables and Spectral Indices

Each annual predictor stack contained six surface reflectance bands (Blue, Green, Red, NIR, SWIR1, and SWIR2) and six spectral indices selected to improve discrimination among water, built-up land, agriculture, and forest. Normalized Difference Vegetation Index (NDVI) captures vegetation greenness, Normalized Difference Water Index (NDWI) and Modified Normalized Difference Water Index (MNDWI) enhance open-water signals [19,20], Automated Water Extraction Index ($AWEI_{nsh}$) enhances separation of water from non-water surfaces using the non-shadow variant [21], NDBI supports built-up discrimination [22], and BSI represents exposed/bare soil conditions [23], shown in Table 1. Predictor redundancy was assessed using Spearman correlation matrices calculated separately from each year's Development samples. Absolute correlations were averaged across the 26 years. Descriptive predictor importance was obtained from Development-fitted RF

models using mean decrease in impurity and from XGBoost models using normalized total gain. These model-specific importance measures were not interpreted causally or used to select the reduced predictor set.

Table 1. Remote sensing predictor variables used for LULC classification.

Predictor Group/Index	Variables or Formula	Primary Role
Spectral bands	Blue, Green, Red, NIR, SWIR1, SWIR2	Surface reflectance/spectral shape
NDVI	$(NIR - Red)/(NIR + Red)$	Vegetation greenness [24]
NDWI	$(Green - NIR)/(Green + NIR)$	Open-water enhancement [19]
MNDWI	$(Green - SWIR1)/(Green + SWIR1)$	Water/Built-up separation [20]
$AWEI_{nsh}$	$4(Green - SWIR1) - (0.25NIR + 2.75SWIR2)$	Water extraction under non-shadow conditions [21]
NDBI	$(SWIR1 - NIR)/(SWIR1 + NIR)$	Built-up response [22]
BSI	$[(SWIR1 + Red) - (NIR + Blue)]/[(SWIR1 + Red) + (NIR + Blue)]$	Bare/exposed soil response [23]

Correlated predictors can influence the classifiers differently. The RBF SVM uses distances in the standardized predictor space, so several correlated variables can give greater weight to shared spectral information. Standardization adjusts scale but does not remove this redundancy. Tree-based models select individual splitting variables, and correlated predictors can substitute for one another or share importance. We therefore used the predefined index ablation experiment to assess the stability of the model comparison empirically; retaining correlated predictors was not assumed to affect all classifiers equally.

A predefined index ablation analysis compared the original twelve predictors with the six reflectance bands alone. Each reduced classifier was tuned separately using the original hyperparameter grid and the same five frozen Development folds. SVM scaling remained within the fitting pipeline. The twelve-predictor models were refitted with their cross-validation-selected parameters, and their Test predictions were confirmed to be identical to those reported in Section 3.1. Both configurations were evaluated on the unchanged outer Test samples. The joint spatial cluster bootstrap described in Section 2.6 was extended across classifiers and predictor configurations using identical group multiplicities. Removing the indices was not assumed to eliminate correlation among the remaining bands.

2.4. Reference Samples, Quality Control, and Spatial Validation Design

Reference samples were prepared by visual interpretation of historical satellite imagery in Google Earth Pro. A total of 1000 point-based reference samples were compiled for each year from 2000 to 2025, yielding 26,000 point year reference records. The same point locations and identifiers were maintained throughout the study period. For each year, the land use/land cover (LULC) class at every reference location was checked against the corresponding historical imagery. Labels were updated whenever a change in LULC was identified and retained when the interpreted class remained unchanged. Labels were assigned first for 2025, when high-resolution imagery covered the whole study area, and then checked year by year at the same locations, back to 2000; a label was changed only where the imagery showed a different land cover. Because each year started from the label of the following year, this procedure may favor retaining labels, so some changes may be under-recorded (Section 4.4). Sampling was purposive: reference points were placed at clearly identifiable locations using pins rather than delineating class polygons.

Winter imagery from November to February was preferred because the ground surface could generally be interpreted more clearly. When winter imagery was unavailable or insufficiently clear, summer imagery from March to June was used. For most of 2000–2010,

high-resolution imagery in Google Earth Pro did not cover the whole study area. Where it was missing for a location and year, imagery from the adjacent years was examined to confirm whether the label should be retained or changed. Only the reference year of each label was recorded, not the acquisition date or the image source. Locations with mixed or unclear class boundaries were avoided as far as practicable. No numerical image purity threshold was applied. Each reference point was associated with the raster pixel containing its coordinates for predictor extraction. The point-based interpretation protocol did not use a polygon-based minimum mapping unit.

Four LULC classes were distinguished: Water, Built-up, Agriculture and Forest. Water included visible lake surface water, canal water and exposed water between or around floating gardens. Agriculture included paddy fields, other cropland, fallow fields, bare land, orchards, swidden fields and cultivated floating garden beds. Built-up included houses, buildings, roads, residential areas and urban surfaces. Forest included forests, degraded forests and hilly evergreen forests. Cultivated floating beds were assigned to Agriculture, whereas exposed water separating or surrounding those beds was assigned to Water. These operational definitions are summarized in Table 2.

Table 2. Operational definitions used for the four land cover classes.

Code	Class	Included Surfaces and Interpretation Rule
1	Water	Visible lake surface water, canal water, and exposed water between or around floating gardens.
2	Built-up	Houses, buildings, roads, residential areas and urban surfaces. Bare land was labelled Agriculture; a separate bare soil class was not used.
3	Agriculture	Agricultural land: paddy fields, other cropland, fallow fields, bare land, orchards, swidden fields and cultivated floating garden beds. Cultivated floating beds are labelled Agriculture.
4	Forest	Forest, degraded forest and hilly evergreen forest as interpreted in the reference imagery.

Technical quality control was performed before model development. Input checks required complete annual records, recognized class labels, finite geographic coordinates, unique point identifiers and coordinates within each year, and consistent identifiers and locations across years. Raster checks excluded points outside the raster extent, points with any masked or NoData predictor and points with any nonfinite predictor value. If multiple records occupied the same valid raster pixel and had conflicting labels, all conflicting records were excluded. Where their labels agreed, one record was retained according to the predefined record priority rule. The quality control criteria and implementation thresholds are provided in Appendix A.1.

Final technical quality control retained 25,948 of the 26,000-candidate point year records (99.80%) and excluded 52 records (0.20%). Each annual dataset contained 998 valid records, comprising 700 Development and 298 Test records. Over the 26 years, 18,200 records were assigned to Development and 7748 to Test. The outer spatial group assignment was fixed before model development and was not changed during quality control. Annual quality control totals and class-specific sample counts are reported in Appendix A. The technical rejection rate does not measure reference label interpretation accuracy.

Figure 1c displays the 2025 class labels at the 998 retained reference locations, while Figure 1d shows their fixed Development/Test assignment and the occupied physical blocks. The same locations and outer split were retained across 2000–2025; the displayed 2025 labels therefore do not represent the labels of every year.

For spatial validation, coordinates were projected to UTM Zone 47N (EPSG:32647) and assigned to a fixed grid of 1500 m square blocks using floor (Easting/1500) and floor (Northing/1500). The reference location pool occupied 449 physical blocks. Blocks connected by predefined point pair identifiers or shared raster pixels were merged into linked groups, producing 381 master Split Groups. Linked groups could contain more than one physical block. The 1.5 km block width was a fixed design choice, rather than an estimated spatial autocorrelation range, and no additional buffer was imposed between neighboring blocks.

An approximately 70% Development and 30% Test partition was selected from 500 GroupShuffleSplit candidates using random seed 42. Selection considered class representation and balance relative to the target Test fraction, without using model performance. The master outer assignment was fixed across years and classifiers. Each final annual dataset represented 449 physical blocks and 381 linked Split Groups, comprising 266 Development groups and 115 Test groups. Across all years, 18,200 records were assigned to Development and 7748 to Test. Integrity checks confirmed no overlap of point identifiers, predefined point pairs, physical blocks, linked groups or raster pixels between the subsets within each year. Grouped five-fold cross-validation was used within Development every year, without a non-grouped fallback, as described in Section 2.5. These internal evaluations share purposively selected spatial support across years and are separate from the probability-based production map assessment.

2.5. Machine Learning Classification and Hyperparameter Tuning

Three supervised classifiers were compared. RF was used because it is a robust tree ensemble benchmark that is commonly used for multidimensional remote sensing classification [13]. SVM with a radial basis function kernel was included because of its strong generalization performance in nonlinear, moderate sample remote sensing problems [15]. XGBoost was included as a regularized gradient-boosting approach capable of representing complex nonlinear relationships [14]. Classification and evaluation were implemented in Python using scikit-learn and XGBoost [14,25].

Hyperparameters were tuned separately for each year and classifier by grid search (Table 3), maximising mean Development-only cross-validation macro-F1. Five-fold StratifiedGroupKFold assignments were generated for each annual Development subset using Split_Group as the grouping variable, shuffle = True and random_state = 42. The assignments were generated with scikit-learn 1.8.0, saved explicitly and reused for RF, SVM and XGBoost, rather than constructed under the model-fitting software version. All 130 year-specific folds contained all four classes in both training and validation subsets and passed checks for point, pair, block, group and raster pixel separation. Five-fold grouped validation was feasible and used every year; no non-grouped fallback was applied.

Table 3. Machine learning hyperparameter used for annual classification.

Model	Hyperparameters	Implementation
RF	n_estimators: 300, 500; max_depth: None, 10, 20; max_features: sqrt, 0.5; min_samples_leaf: 1, 2	scikit-learn
SVM	RBF kernel; C: 1, 10, 100; gamma: scale, 0.01, 0.1; StandardScaler pipeline	scikit-learn
XGBoost	n_estimators: 200, 400; max_depth: 3, 5, 7; learning_rate: 0.03, 0.10; subsample: 0.8; colsample_bytree: 0.8	XGBoost

SVM was implemented as a StandardScaler-RBF SVC pipeline, so scaling was fitted only to the training records within each fold and subsequently to the full Development subset during refitting. RF and XGBoost used unscaled predictors. All classifiers used

the same 12 spectral predictors, with identifiers, coordinates, years, labels and split/group fields excluded from the predictor matrix. After tuning, each selected model was refitted to its complete annual Development subset and evaluated on the unchanged outer Test subset. This produced 78 annual model evaluations (three classifiers × 26 years), fitted with scikit-learn 1.6.1 and XGBoost 3.4.1. The SVM run used probability = False and generated hard class predictions.

2.6. Accuracy Assessment, Uncertainty, and Model Selection

For each annual model evaluation, uncertainty in Test classification metrics was summarized using 1000 percentile cluster bootstrap replicates. The resampling unit was the held-out Split Group. Each replicate sampled the same number of Test groups with replacement and retained all records in each selected group, including multiplicities. The 2.5th and 97.5th percentiles formed the reported 95% intervals. Identical group draws were used across classifiers within a year. These intervals describe resampling variability conditional on the purposively selected reference locations and fitted models; they do not quantify probability-based watershed-wide accuracy, reference label error, or class area and area change uncertainty. Twenty-six-year performance summaries are unweighted means of the annual metrics and are distinct from metrics pooled over all point year records [26–29].

Classifier performance was summarized using unweighted means of the 26 annual internal Test metrics. The classifier used for production mapping was selected by the highest mean Development-only grouped cross-validation macro-F1 across the 26 years; outer Test scores were excluded from this selection.

Pairwise differences in mean annual OA and macro-F1 were assessed using a joint spatial cluster bootstrap. The resampling pool comprised the 115 unique master Test Split Groups represented across the study period. Each of 2000 replicates sampled 115 groups with replacement, using random seed 42. Identical group multiplicities were applied to every classifier and year, retaining all available records within each selected group. Annual metrics were recomputed before averaging across the 26 years, and paired classifier differences were calculated within each replicate. The 2.5th and 97.5th percentiles formed marginal 95% intervals, without adjustment for multiple comparisons. Years were retained as a fixed repeated panel rather than resampled independently. These intervals are conditional on the fitted models and purposively selected Test group pool; they exclude model refitting and reference label uncertainty and do not account for residual dependence between different spatial groups.

In the following equations, n_{ii} denotes the number of correctly classified internal Test samples in class i, N is the total number of internal Test samples, and k = 4 is the number of classes. UA_i is the proportion of samples assigned to class i that are correctly classified, whereas PA_i is the proportion of reference samples belonging to class i that are correctly classified. OA and class-specific F1 are calculated as proportions and expressed as percentages where indicated. Macro-F1 is the unweighted mean of the four class-specific F1-scores. These count-based equations describe internal Test evaluation; probability-based production map accuracy is calculated from the area-weighted error matrix described below.

$$OA = \frac{\sum_{i=1}^{k} n_{ii}}{N}$$

$$F1_i = \frac{2 \times UA_i \times PA_i}{UA_i + PA_i}$$

A stratified probability reference panel was used to assess the fixed production maps for 2000, 2010, 2020 and 2025. A total of 600 raster grid pixels were selected by simple random sampling without replacement within 16 mapped 2000–2025 transition strata. The

same sampled pixels formed the four-year reference panel, and their inclusion probabilities and stratum population sizes were retained. An overlap check identified one sampled pixel that also occurred in the approved reference pool used to fit the production models. This pixel was retained to preserve the original probability sample. The production map assessment is therefore separate from the internal spatially grouped Test evaluation but is not completely independent of the production model training data.

The 16 strata represented the combinations of the four mapped classes in 2000 and 2025. Sample allocation assigned a minimum of 20 pixels to each nonempty stratum, with the remaining allocation distributed in proportion to stratum pixel populations. Random selection used seed 20261020 and was conducted within the fixed common valid domain without spatial displacement or minimum-distance filtering. The total of 600 locations represented a feasible interpretation workload rather than a sample size derived from a prespecified precision target. The sampling uncertainty achieved by this design is reported with the resulting area and net change estimates.

For the 2000 reference assessment, interpretation used available historical imagery from within the calendar year. The reference records document the year rather than verified day- and month-specific acquisition dates; temporal provenance is therefore reported at year precision. Differences in acquisition timing within the year may introduce seasonal interpretation uncertainty, which is not included in the reported sampling intervals.

Reference interpretation was undertaken by the three authors, with responsibility allocated by year. Initial reference labels were assigned without viewing the SVM predictions. The resulting labels were subsequently cross-checked within the author team to produce the final reference dataset. This process constituted author cross-checking rather than a formal independent repeat interpretation experiment. Only the final labels were retained; separate interpreter-specific label sets and records of label changes during cross-checking were not preserved. Each label was assigned to a confidence level (High, Medium or Low) at the time of interpretation.

For each class and reference year, area was estimated as the sum, across strata, of the stratum pixel population multiplied by the sample mean of pixel area times the reference class indicator. Sampling variances used the stratified finite population estimator [30]. Net changes were estimated from paired class indicator differences at the same pixels, preserving temporal covariance. Approximate 95% sampling intervals were calculated as the estimate plus or minus 1.96 sampling standard errors. These marginal intervals are conditional on reference validity and exclude interpretation error. Map accuracy was estimated from the area-weighted error matrix, with mapped classes in rows and reference classes in columns. Overall accuracy, Kappa, UA, PA and F1 were reported separately from the internal Test scores. Sensitivity to Low-confidence reference labels was assessed while retaining the original sample and sampling weights. High- and Medium-confidence labels were held fixed, whereas each Low-confidence sample year label was allowed to take any of the four classes. For each class separately, the minimum and maximum reference-based net changes were calculated over the permitted label assignments. These deterministic sensitivity bounds describe the potential effect of the flagged labels; they are not confidence intervals and do not incorporate sampling variability or possible errors in labels assigned higher confidence.

2.7. Annual LULC Mapping and Surface Water Area Estimation

SVM was selected for production mapping using mean Development-only grouped cross-validation performance across the 26 years. For each year, the production SVM was refitted to all 998 approved annual reference records using the hyperparameters selected within Development, then applied to the complete annual predictor raster. These produc-

tion models were distinct from the Development-fitted models used for internal outer-Test evaluation. Production map accuracy was assessed separately using the probability reference panel described in Section 2.6. The same classifier family was used throughout the record, with year-specific fitting and hyperparameters.

Annual raw mapped class areas were calculated from the SVM production maps within the intersection of valid classified pixels across all 26 years. This fixed common domain contained 6,483,344 pixels and covered 5438.20 km^2. For each year and class, area was calculated by summing the actual areas of pixels assigned to that class, accounting for spatial variation in pixel area. Using identical spatial support ensured comparability between annual class area totals and the water transition comparisons. These values are unadjusted classified map totals and exclude areas outside the common valid domain. Reference-based class area and net change estimates, with their sampling uncertainty, were calculated separately using the probability-sampled reference panel.

Within the fixed common domain, every classified pixel belongs to exactly one of the four mutually exclusive classes. Their summed mapped area therefore remains constant at 5438.20 km^2 by construction. A change from Water to Agriculture redistributes area between classes without changing the geographic extent of the analysis domain. The constant total does not imply environmental stability or independently demonstrate classification accuracy. Areas were calculated from the geographic grid cell areas rather than by assuming exactly 900 m^2 per pixel.

Annual trends were assessed using the Mann–Kendall test with tie-corrected variance, a continuity correction of one unit, $Z = (S - \mathrm{sgn}(S))/\sqrt{\mathrm{Var}(S)}$, because S changes in steps of two, and two-sided normal approximation p-values. The Mann–Kendall test was implemented using a custom Python routine based on NumPy, with two-sided normal tail probabilities calculated using the standard library math.erfc function. A nominal significance threshold of $\alpha = 0.05$ was used. Sen's slope was calculated as the median of all pairwise annual slopes. Nominal 95% confidence intervals for Sen's slope were calculated from the ordered pairwise slopes using the rank-based normal approximation implemented in scipy.stats.theilslopes (SciPy 1.17.0). The interquartile range was calculated as the 75th minus the 25th percentile of the 325 pairwise slopes, using linear percentile interpolation. These intervals were not adjusted for serial dependence or classification and reference label uncertainty.

Sensitivity to lag-1 dependence was examined using trend-free pre-whitening: the original Sen trend was removed, lag-1 autocorrelation was estimated from the mean-centered detrended series, the previous detrended value multiplied by this coefficient was subtracted, and the original trend was restored. The Mann–Kendall test was then applied to the resulting 25 annual values. Reported trend p-values are nominal and unadjusted for multiple testing.

For each class, an additive area correction was calculated as the reference-based area estimate minus the raw mapped area in 2000, 2010, 2020 and 2025. Corrections were interpolated linearly between these four reference years. A scenario adding the full interpolated correction to the raw annual series was compared with the unadjusted series. These scenarios represent assumed annual correction patterns rather than independently validated annual areas or confidence intervals. Their trend tests do not propagate sampling uncertainty in the reference year corrections or reference interpretation error. To examine sensitivity to the analysis period and avoid crossing a sensor transition within the tested series, trends were also assessed separately for the Landsat 8 OLI-only sub-record, 2013–2025. The same fixed common domain and raw mapped class-area series were used. Mann–Kendall tests and Sen's slopes were calculated from 13 annual values, and trend-free pre-whitening produced

12 values for the corresponding sensitivity tests. This comparison does not isolate sensor effects from differences in observation timing, reference interpretation or classification error.

2.8. Surface Water Persistence, Gain, and Loss Analysis

Each annual SVM land cover map was converted to a binary water mask, assigning Water a value of 1 and all other classes a value of 0. Pairs of masks were compared within the fixed common valid domain to identify persistence (Water at both endpoints), gain (non-Water changing to Water) and loss (Water changing to non-Water). Areas were calculated by summing the actual areas of the corresponding pixels, and net mapped change was calculated as gain minus loss. Comparisons covered 2000–2010, 2010–2020, 2020–2025 and 2000–2025, matching the four reference years used for paired reference-based change estimation. Persistence denotes agreement between the two endpoint maps and does not imply uninterrupted water cover during the intervening years.

2.9. Decomposition of the Water Class

To separate Inle lake from other water bodies, the common domain was divided into four fixed zones that were applied to every year. The lake zone is the Inle lake open-water polygon (Forest Department, Myanmar), which covers 45.40 km^2 of the common domain. The other zones were derived from the boundary of the Inle (Inlay) Wetland Bird Sanctuary, gazetted in 1985 (boundary from the Forest Department, Myanmar). The part of the sanctuary south of 20.02°N, the narrowest point of the Balu Chaung corridor, contains the Moebye reservoir basin and forms the reservoir zone. The rest of the sanctuary, excluding the lake polygon, forms the wetland zone; it comprises lake-margin marsh, floating gardens, seasonally flooded fields and the Balu Chaung corridor. The domain outside the sanctuary forms the other zone, which contains ponds, channels and isolated Water pixels. Annual mapped Water area was summed within each zone and tested for trends with the procedures of Section 2.7, for 2000–2025 and for the Landsat 8-only period. Reference-based Water changes for each zone were estimated by restricting the stratified estimator (Section 2.6) to sample pixels within that zone (domain estimation); because the sample was not stratified by zone, these estimates are imprecise.

3. Results

3.1. Classification Accuracy and Model Comparison

Across the 26 annual spatially grouped internal Test evaluations, SVM achieved the highest mean overall accuracy (71.49%), compared with RF (68.40%) and XGBoost (68.33%) (Table 4; Figure 3). SVM also achieved the highest mean kappa (0.620) and macro-F1 (71.79%), compared with 0.578 and 68.87% for RF, and 0.577 and 68.73% for XGBoost, respectively. These summaries are unweighted means of the annual internal Test metrics, conditional on the purposively selected reference locations. They are distinct from the area-weighted accuracy estimates obtained from the four-year probability-based assessment of the production maps.

Paired cluster bootstrap comparisons used 2000 replicates of the 115 master Test groups, with identical group multiplicities applied across all years and classifiers. SVM exceeded RF in mean annual OA by 3.08 percentage points (95% marginal percentile interval: 1.76–4.61) and XGBoost by 3.16 points (2.03–4.47). Corresponding mean macro-F1 differences were 2.92 points (1.65–4.24) and 3.06 points (1.99–4.20), respectively. The RF-XGBoost intervals included zero for both metrics. These intervals describe conditional resampling variability for the fixed fitted models and purposively selected Test group pool; they are not probability-based watershed-wide accuracy intervals and are not adjusted for multiple comparisons.

Table 4. Twenty-six-year average classifier performance.

Model	Mean OA (%)	Mean Kappa	Mean Macro-UA (%)	Mean Macro-PA (%)	Mean Macro-F1 (%)
RF	68.40	0.578	69.46	68.82	68.87
SVM	71.49	0.620	72.20	71.91	71.79
XGBoost	68.33	0.577	69.20	68.78	68.73

Note: Values are unweighted means of 26 annual internal Test metrics. Macro-UA, macro-PA and macro-F1 are unweighted averages across the four classes within each year. These results are separate from the probability-based production map assessment.

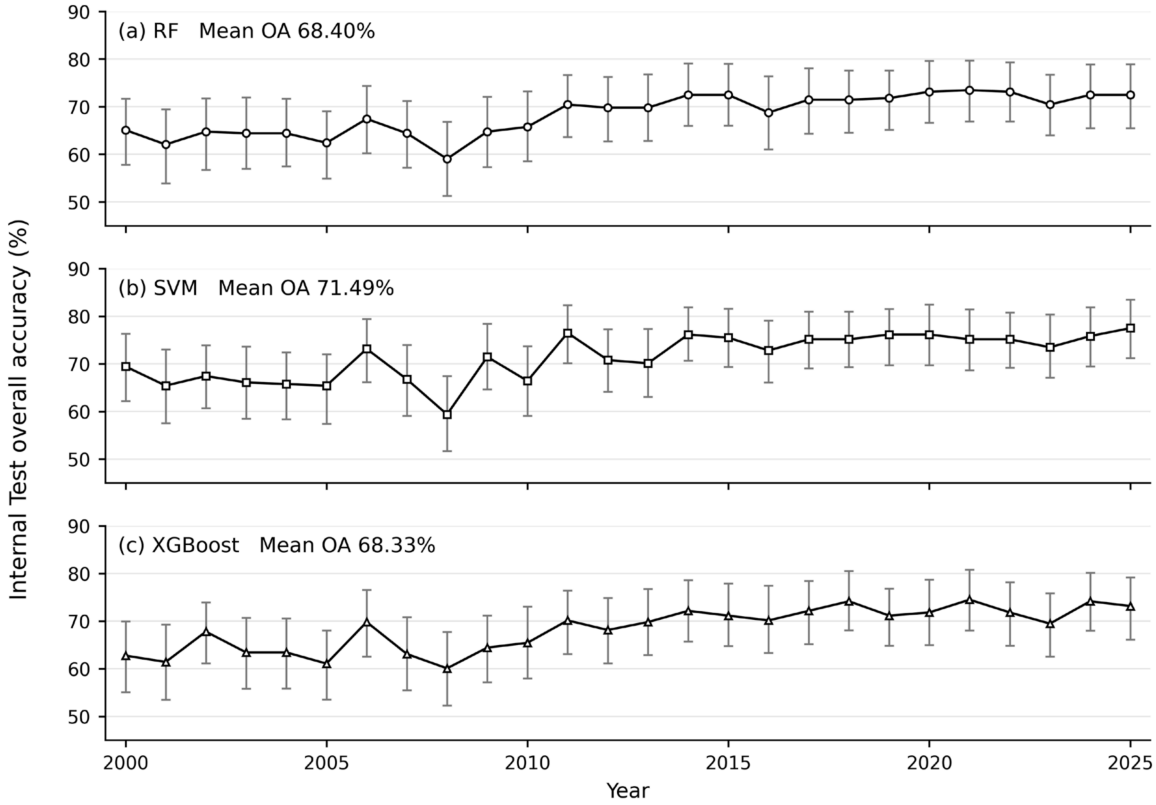

Figure 3. Annual internal Test overall accuracy of (**a**) RF, (**b**) SVM and (**c**) XGBoost for 2000–2025. Symbols show annual accuracy estimates, connecting lines link successive years, and error bars show pointwise 95% percentile intervals from 1000 cluster bootstrap replicates of the held-out Split Groups. Panel means are unweighted averages of the 26 annual accuracies. The panels use identical axis scales. Intervals are conditional on the purposively selected Test locations and fitted models; they exclude reference label and model refitting uncertainty and do not represent probability-based watershed-wide accuracy intervals or simultaneous coverage across years and classifiers.

Class-specific SVM performance varied across the four classes in the internal Test evaluation (Table 5). Forest achieved the highest mean annual F1-score (85.79%), followed by Water (75.19%), Built-up (68.67%) and Agriculture (57.53%). Agriculture had mean UA and PA of 58.02% and 57.48%, respectively, indicating substantial commission and omission errors within the sampled Test locations. Water had mean UA of 80.09% and PA of 71.36%, showing that its internal classification performance also remained imperfect. These unweighted annual internal Test summaries are distinct from the area-weighted production-map assessment reported in Table 6 and Figure 4.

Table 5. Mean class-specific SVM accuracy across 2000–25.

Class	Mean UA (%)	Mean PA (%)	Mean F1 (%)
Water	80.09	71.36	75.19
Built-up	66.51	71.15	68.67
Agriculture	58.02	57.48	57.53
Forest	84.17	87.64	85.79

Note: Values are unweighted means of the 26 annual internal Test metrics. Mean F1 is the average of annual class-specific F1-scores, rather than the harmonic mean of the displayed mean UA and mean PA. These results are separate from the probability-based production map assessment.

Table 6. Area-weighted accuracy of the SVM production maps for four reference years.

Year	N	Weighted OA %	95% Sampling Interval %	Kappa	Macro F1
2000	600	58.7	54.8 to 62.6	0.370	0.527
2010	600	64.3	60.0 to 68.6	0.444	0.592
2020	600	71.4	67.3 to 75.5	0.525	0.653
2025	600	69.8	66.1 to 73.6	0.512	0.640

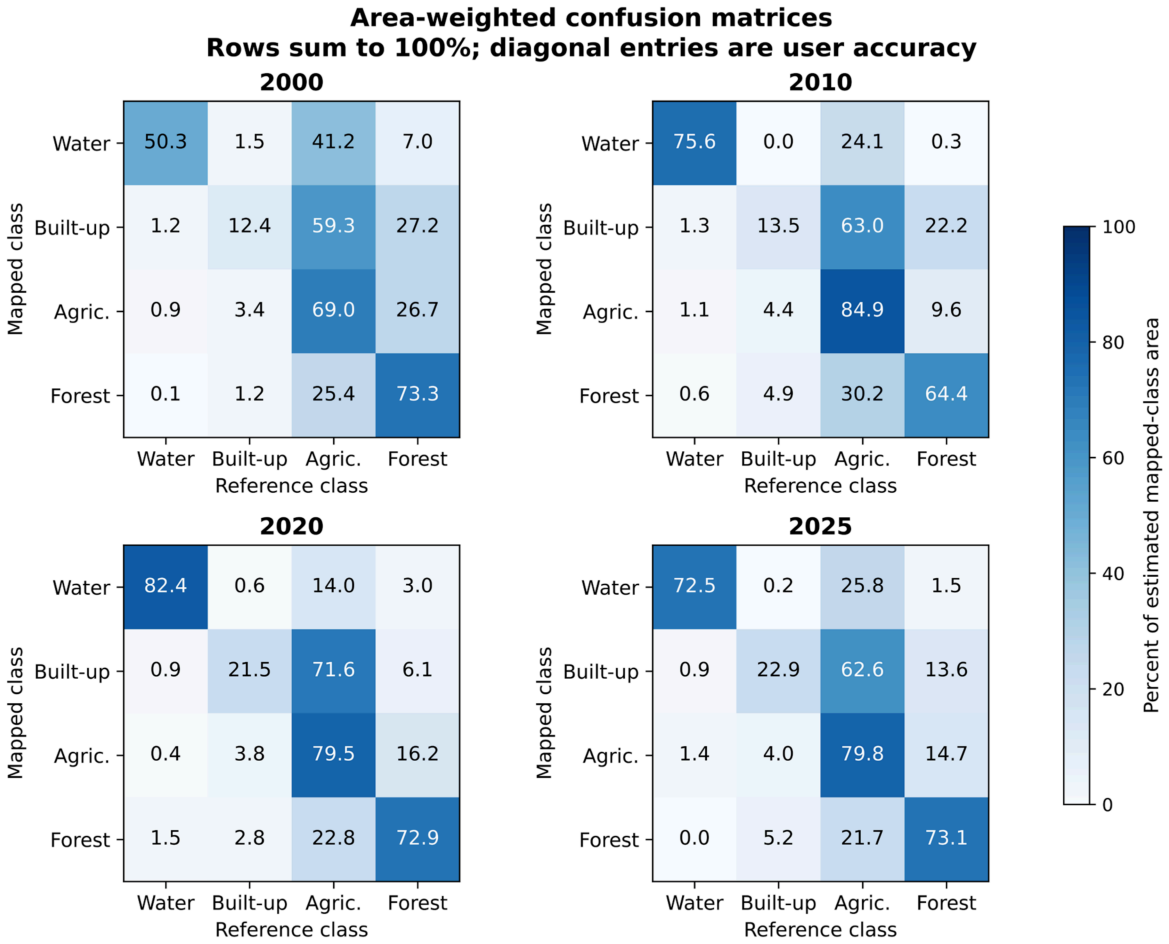

Figure 4. Area-weighted confusion matrices for the SVM production maps in 2000, 2010, 2020 and 2025. Rows represent mapped classes and columns represent reference classes. Values are row percentages; diagonal entries indicate user accuracy.

Area-weighted overall accuracy was 58.7%, 64.3%, 71.4% and 69.8% in 2000, 2010, 2020 and 2025, respectively. Built-up user accuracy remained low at 12.4–22.9%.

The predictor diagnostics identified substantial redundancy. Mean absolute within-year Development Spearman correlations were 0.995 between SWIR1 and $AWEI_{nsh}$, 0.978 between SWIR2 and $AWEI_{nsh}$, and 0.975 between SWIR1 and SWIR2. MNDWI had the highest mean normalized importance in both RF (0.172) and XGBoost (0.212), followed by NDWI (0.125 and 0.154, respectively). These descriptive importances reflect the fitted models and the sharing of information among correlated predictors. The correlation matrix and predictor importance summaries are presented in Figures A1 and A2.

Using six reflectance bands alone, mean annual Test OA was 68.50% for RF, 71.39% for SVM and 68.22% for XGBoost; corresponding macro-F1 values were 68.85%, 71.62% and 68.58%. SVM exceeded RF by 2.89 percentage points in OA (conditional 95% interval: 1.57–4.31) and XGBoost by 3.16 percentage points (2.01–4.42). Its corresponding macro-F1 advantages were 2.77 percentage points (1.52–4.08) and 3.04 percentage points (1.96–4.17). Changes in SVM's advantage between the twelve-predictor and six-band configurations had intervals including zero. Thus, SVM's advantage persisted in this index ablation comparison, although substantial correlation remained among the six bands.

Selected year internal confusion matrices show interchange between Built-up and Agriculture (Table A2). In 2012, 23 of 76 reference Built-up observations were classified as Agriculture, while 15 of 84 reference Agriculture observations were classified as Built-up. Built-up internal user and producer accuracies were 66.67% and 63.16%, respectively. These results describe the Development-fitted SVM on the purposively selected outer Test samples and do not estimate population-level production map accuracy in 2012. Because bare soil was not a separate reference class, its specific contribution to the observed confusion cannot be quantified from these matrices.

3.2. Annual LULC Maps and Spatial Patterns

The rasterized watershed contained 6,485,051 pixels, covering 5439.64 km^2. Annual valid coverage was complete within this boundary in 2000 and 2014. In 2013, 1705 pixels were invalid, representing 0.0263% of watershed pixels and 1.434 km^2. The intersection of valid classified pixels across all 26 years retained 6,483,344 pixels (5438.20 km^2), excluding 1707 watershed pixels in total. Within this common domain, all pixels were classified in 2000, 2013 and 2014, so the NoData fraction was zero for each of these years. Consequently, annual mapped area and transition comparisons use identical spatial coverage. Complete coverage does not establish classification accuracy.

The annual classified maps make it possible to examine change in land cover not only as a time series of areas but also as a spatial pattern across the watershed. Representative SVM-derived maps for 2000, 2005, 2010, 2015, 2020, and 2025 show a landscape organized primarily by topography: Forest persists mainly in the peripheral uplands and hill slope zones ringing the watershed, while agriculture occupies the broad valley floor and lowland areas around the lake (Figure 5). The representative maps depict differences in the spatial distribution of Agriculture and Forest; these patterns do not independently establish Forest-to-Agriculture conversion.

Built-up land follows a markedly different spatial logic from either forest or agriculture. Rather than form a compact urban footprint, built-up pixels are scattered throughout the agricultural lowlands and frequently intermixed with agriculture and exposed soil. This spatial pattern is consistent with the classification accuracy results in Section 3.1, wherein Built-up registered a lower mean F1-score than forest or water. This implies that a class that is spatially and spectrally entangled with its surroundings is inherently harder to classify consistently from year to year; hence, its area series is interpreted with caution in the quantitative comparisons that follow rather than being read as a literal settlement record.

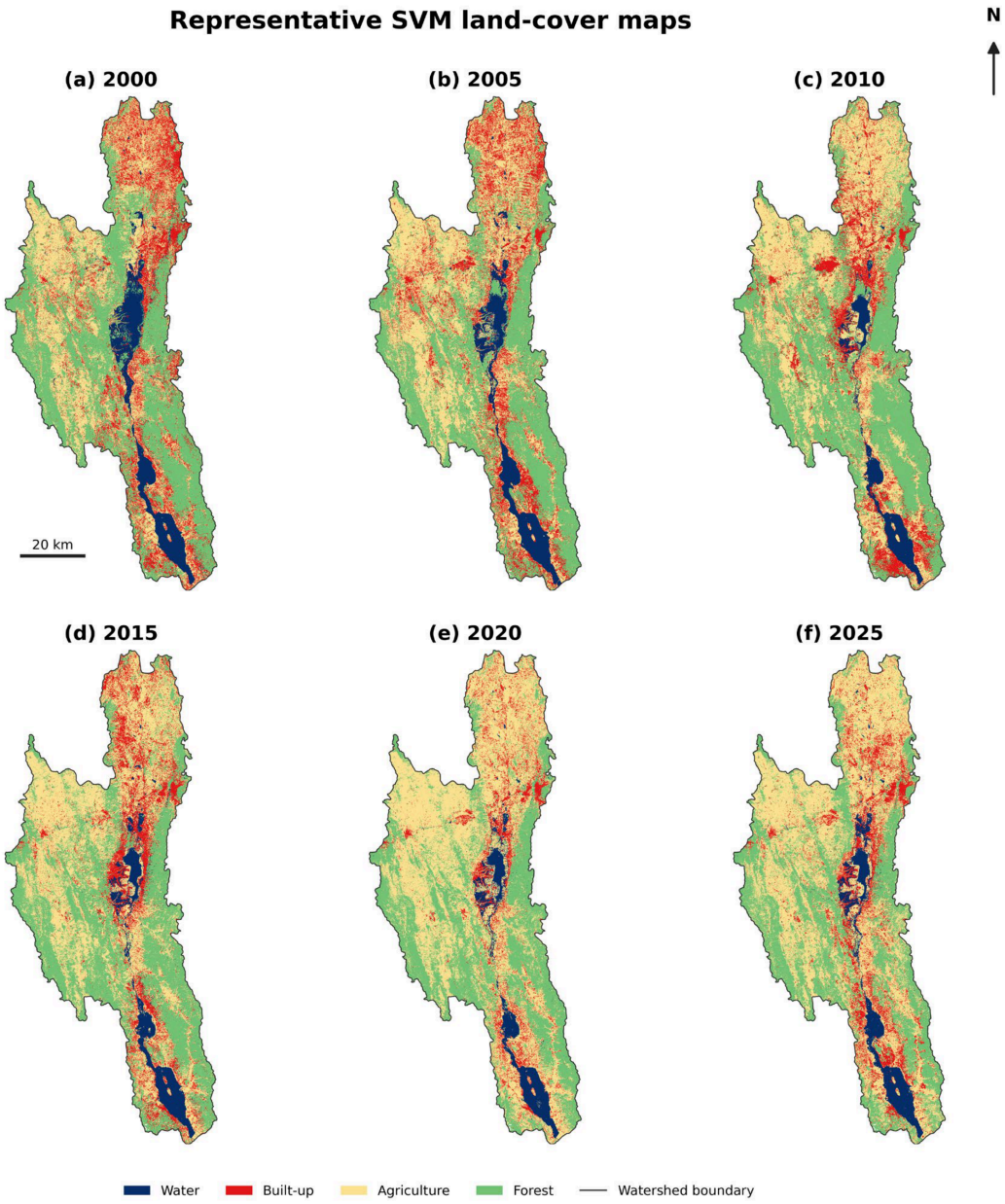

Figure 5. Representative SVM land cover maps for 2000, 2005, 2010, 2015, 2020 and 2025 within the fixed common valid domain of 6,483,344 pixels (5438.20 km^2). Maps use WGS84 geographic coordinates (EPSG:4326) with north-up orientation. All panels share the same extent and display scale; the common legend, 20 km scale bar and north arrow apply to every panel. The outline shows the rasterized watershed boundary, and pixels outside the common domain are blank.

3.3. Annual LULC Area Trajectories

The annual SVM mapped area series (Figure 6) was evaluated over a fixed common domain of 6,483,344 pixels (5438.20 km^2). Raw mapped surface water area decreased from 447.36 km^2 in 2000 to 349.15 km^2 in 2025, a net difference of −98.20 km^2 (−21.95%). Across the 26 annual maps, the minimum and maximum mapped water areas occurred in 2014 (268.63 km^2) and 2000 (447.36 km^2), respectively. These values describe unadjusted mapped surface water across the watershed. Reference-based net change estimates and their sampling uncertainty are reported separately in Section 3.4.

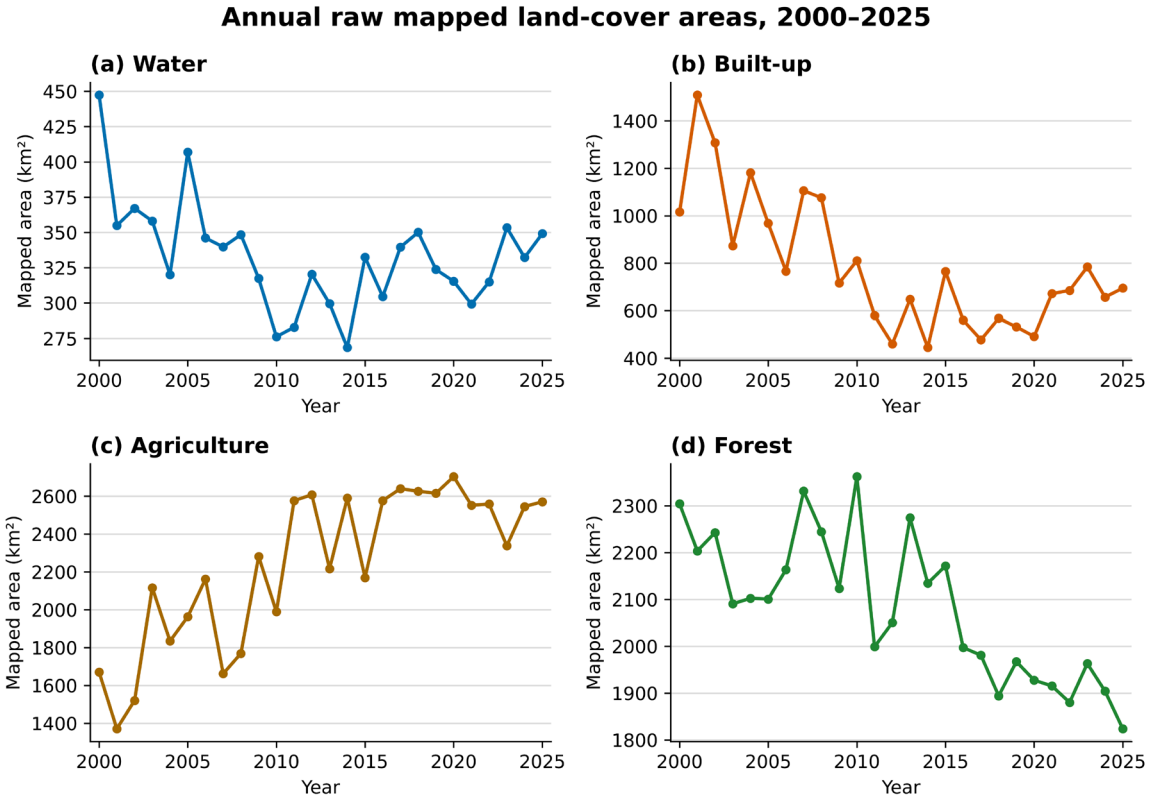

Figure 6. Annual raw mapped areas of Water, Built-up, Agriculture and Forest within the fixed common domain of 5438.20 km^2, 2000–2025. Values are unadjusted classified map totals.

Within the same fixed common domain, raw mapped Agriculture increased from 1670.60 km^2 in 2000 to 2570.69 km^2 in 2025, a net increase of 900.10 km^2 (53.88%). Forest decreased from 2304.48 to 1823.77 km^2, a net reduction of 480.71 km^2 (20.86%). Raw mapped Built-up area decreased from 1015.77 to 694.58 km^2, a difference of −321.19 km^2 (−31.62%). These changes describe differences between classified map totals. In particular, the mapped Built-up decline should not be interpreted as settlement contraction because the probability-based assessment identified substantial commission error in this class. Reference-based net change estimates and sampling intervals are presented separately in Section 3.4.

The tie-corrected Mann–Kendall test (Table 7) applied to the 26-year raw mapped area series indicated a positive trend in Agriculture ($\tau = +0.538$, $p = 0.0001$; Sen's slope = +40.13 km^2 yr^{-1}) and negative trends in Built-up ($\tau = -0.446$, $p = 0.0015$; Sen's slope = −22.83 km^2 yr^{-1}) and Forest ($\tau = -0.594$, $p < 0.0001$; Sen's slope = −15.63 km^2 yr^{-1}). Water showed a negative trend that did not reach the nominal 0.05 significance level ($\tau = -0.255$, $p = 0.0707$; Sen's slope = −1.65 km^2 yr^{-1}). These results describe the unadjusted mapped area trajectories, and the reported p-values are not adjusted for serial dependence or multiple testing. Sensitivity to lag-1 serial dependence and assumed classification error corrections is reported in Table 8.

Agriculture and Built-up significance changes under the assumed correction patterns. Forest remains below 0.05 in these specific tests. Water changes slope sign and can appear significant under the full imposed correction. That apparent scenario significance does not establish a real Water increase: the endpoint sampling interval includes zero, and intermediate-year corrections have not been observed. Within the Landsat 8-only sub-record (2013–2025), Forest retained a negative mapped trend, with Sen's slope of −24.89 km^2 yr^{-1} (Mann–Kendall $p = 0.0005$; trend-free pre-whitening $p = 0.0020$). Water, Built-up and Agriculture had slopes of +3.22, +14.20 and −1.21 km^2 yr^{-1}, respectively, with corresponding Mann–Kendall p-values of 0.1606, 0.0995 and 0.9514 and pre-whitening p-values of 0.2437,

0.0641 and 0.6312. Although mapped Agriculture was 353.81 km^2 greater in 2025 than in 2013, its annual trajectory did not show a significant monotonic increase. These results demonstrate sensitivity to the analysis period and support caution when interpreting the full multi-sensor record.

Table 7. Mann–Kendall trend statistics for the annual raw mapped area series, 2000–2025.

Class	S	Kendall's τ	Z	*p*-Value	Sen's Slope (km^2 yr^{-1})	Nominal 95% Slope CI (km^2 yr^{-1})	Pairwise Slope IQR (km^2 yr^{-1})
Water	−83	−0.255	−1.807	0.0707	−1.65	−3.88 to +0.17	9.00
Built-up	−145	−0.446	−3.174	0.0015	−22.83	−37.09 to −10.21	52.89
Agriculture	175	+0.538	+3.835	0.0001	+40.13	+24.74 to +56.43	69.38
Forest	−193	−0.594	−4.232	<0.0001	−15.63	−19.27 to −10.93	22.73

Note: No exact ties occurred in the annual mapped-area values; the tie-corrected Var(S) was therefore 2058.33 for each class. Z was calculated as $Z = (S - 1)/\sigma$ for $S > 0$ and $Z = (S + 1)/\sigma$ for $S < 0$, where $\sigma = \sqrt{\mathrm{Var}(S)} = 45.369$ (e.g., Water: $(-83 + 1)/45.369 = -1.807$); *p*-values are two-sided normal probabilities of Z. Lag-1 Pearson correlations of the raw annual series were 0.396, 0.663, 0.701 and 0.619 for Water, Built-up, Agriculture and Forest, respectively. The corresponding detrended lag-1 autocorrelation coefficients used for trend-free pre-whitening were 0.242, 0.336, 0.153 and 0.140. Slope confidence intervals are nominal and unadjusted for serial dependence; the separate pre-whitening sensitivity results are reported in Table 8. The pairwise slope IQR is a descriptive spread, not a confidence interval.

Table 8. Sensitivity of annual trend test *p*-values to assumed classification error corrections. The full correction pattern interpolates area corrections between the four reference years; it does not represent independently validated annual areas.

Class	Raw MK *p*	Raw TFPW *p*	Full Pattern MK *p*	Full Pattern TFPW *p*
Water	0.0707	0.3624	0.0008	0.0012
Built-up	0.0015	0.0016	1.0000	0.8701
Agriculture	0.0001	0.0004	0.3321	0.3624
Forest	<0.0001	<0.0001	0.0017	0.0030

3.4. Major Land Cover Changes

The raw mapped area series showed agricultural expansion and forest reduction between 2000 and 2025 (Table 9), with substantial variation among five-year intervals. From 2000 to 2005, mapped Agriculture increased by 292.25 km^2 and Forest decreased by 203.79 km^2. Among the five-year intervals examined, 2015–2020 showed the largest mapped agricultural increase (+534.08 km^2), accompanied by a forest decrease of 244.20 km^2. From 2020 to 2025, both mapped Agriculture and Forest decreased, by 132.75 and 103.83 km^2, respectively. These interval differences describe changes in raw classified map totals; they do not establish direct class-to-class conversion, persistent land use change or causation.

Table 9. Raw mapped land cover areas and endpoint changes within the fixed common domain, 2000–2025.

Class	2000 (km^2)	2025 (km^2)	Net Change (km^2)	Relative Change (%)	Annualized Net Rate (km^2 yr^{-1})
Water	447.36	349.15	−98.20	−21.95	−3.93
Built-up	1015.77	694.58	−321.19	−31.62	−12.85
Agriculture	1670.60	2570.69	+900.10	+53.88	+36.00
Forest	2304.48	1823.77	−480.71	−20.86	−19.23

Note: Values are unadjusted classified map totals. Relative change is calculated against the 2000 area. Annualized net rate is the endpoint difference divided by 25 years and is distinct from Sen's slope.

The built-up series warrants caution since its endpoint declines. Not only are large annual oscillations inconsistent with a simple monotonic urbanization trajectory; but also, they are likely to reflect, at least partially, the classification interchange among built-up, bare/exposed soil, and agricultural surfaces. For this reason, the built-up result is reported as variability in the mapped class area instead of being interpreted as a direct measure of settlement contraction.

The reference-based 2000–2025 changes were Agriculture +450.1 km^2 (95% sampling interval +203.4 to +696.8), Forest −631.5 km^2 (−857.3 to −405.7), Built-up +139.2 km^2 (+23.5 to +254.9), and Water +42.3 km^2 (−23.5 to +108.0). The Water interval included zero, so the reference assessment did not establish a clear full-period increase or decrease. Likewise, the 2020–2025 Water change of −4.6 km^2 (−43.9 to +34.8) did not establish a recovery. These estimates concern watershed-wide surface water within the fixed common domain.

Of the 2400 assessed sample year records, 31 were assigned Low confidence: seven in 2000, eight in 2010, five in 2020 and eleven in 2025. Allowing these labels to vary freely among the four classes yielded class-specific 2000–2025 net change bounds of −29.5 to +182.7 km^2 for Water, +66.6 to +278.8 km^2 for Built-up, +381.1 to +593.2 km^2 for Agriculture, and −775.5 to −563.3 km^2 for Forest. Thus, the estimated directions of agricultural expansion and forest loss persisted under this restricted sensitivity assessment, whereas Water change could take either sign. These bounds are separate from the sampling intervals in Table 10 and do not assess possible errors in high- or medium-confidence labels.

Table 10. Reference-based net land cover changes from 2000 to 2025, with sampling standard errors and approximate 95% sampling intervals.

Class	Net km^2	Sampling SE km^2	95% Sampling Interval km^2
Water	+42.3	33.5	−23.5 to +108.0
Built-up	+139.2	59.0	+23.5 to +254.9
Agriculture	+450.1	125.9	+203.4 to +696.8
Forest	−631.5	115.2	−857.3 to −405.7

Reference-based class areas for each reference year, with sampling standard errors, and changes for each sub-period are given in Tables A3 and A4. Almost all of the reference-based agricultural increase and forest loss occurred between 2000 and 2010: Agriculture increased by 451 km^2 (95% sampling interval +218 to +685) and Forest decreased by 554 km^2 (−765 to −343). Neither class changed significantly during 2010–2020 or 2020–2025. The reference-based Water area was 254 km^2 (193 to 316) in 2000, compared with 447.36 km^2 mapped; from 2010 onward the mapped Water areas fell within the reference sampling intervals. Interpretation confidence was lower in 2000 than in later years: 188 of the 600 labels for 2000 (31%) were medium confidence, compared with 48 (8%) in 2010.

3.5. Surface Water Persistence, Gain and Loss

Pairwise comparisons of the annual SVM water masks were performed for 2000–2010, 2010–2020, 2020–2025 and the full 2000–2025 period (Table 11; Figures 7 and 8). All comparisons used the same fixed common domain of 6,483,344 pixels (5438.20 km^2), ensuring identical spatial support across periods and consistency with the annual mapped area series. Persistence, gain and loss describe transitions in raw mapped watershed-wide surface water. These mapped transitions are distinguished from the reference-based net change estimates and sampling intervals reported in Section 3.4.

Table 11. Reconciliation of raw mapped Water areas and endpoint transitions within the fixed common domain of 5438.20 km^2. All areas are in km^2.

Period	Start-Year Water	End-Year Water	Persistent Water	Gain	Loss	Net Change (Gain − Loss)	Endpoint Difference (End − Start)
2000–2010	447.356	276.307	267.525	8.782	179.831	−171.049	−171.049
2010–2020	276.307	315.398	249.187	66.211	27.120	39.091	39.091
2020–2025	315.398	349.155	274.910	74.245	40.488	33.757	33.757
Sum of sub-periods	—	—	—	149.237	247.438	−98.201	—
2000–2025	447.356	349.155	315.896	33.259	131.460	−98.201	−98.201

Note: Persistence denotes Water at both endpoints, without implying uninterrupted water cover. Persistence + loss equals start year Water; persistence + gain equals end-year Water. Net change equals both gain − loss and end year − start year Water. The sum of the three sub-period net changes equals the full-period endpoint difference. Summed sub-period gains and losses count transitions across successive intervals and can include repeated changes at the same locations; they are not equivalent to full-period endpoint gains and losses. Values are rounded to three decimals; identities were checked using unrounded values. These are raw mapped areas, not reference-adjusted estimates or evidence of transition accuracy.

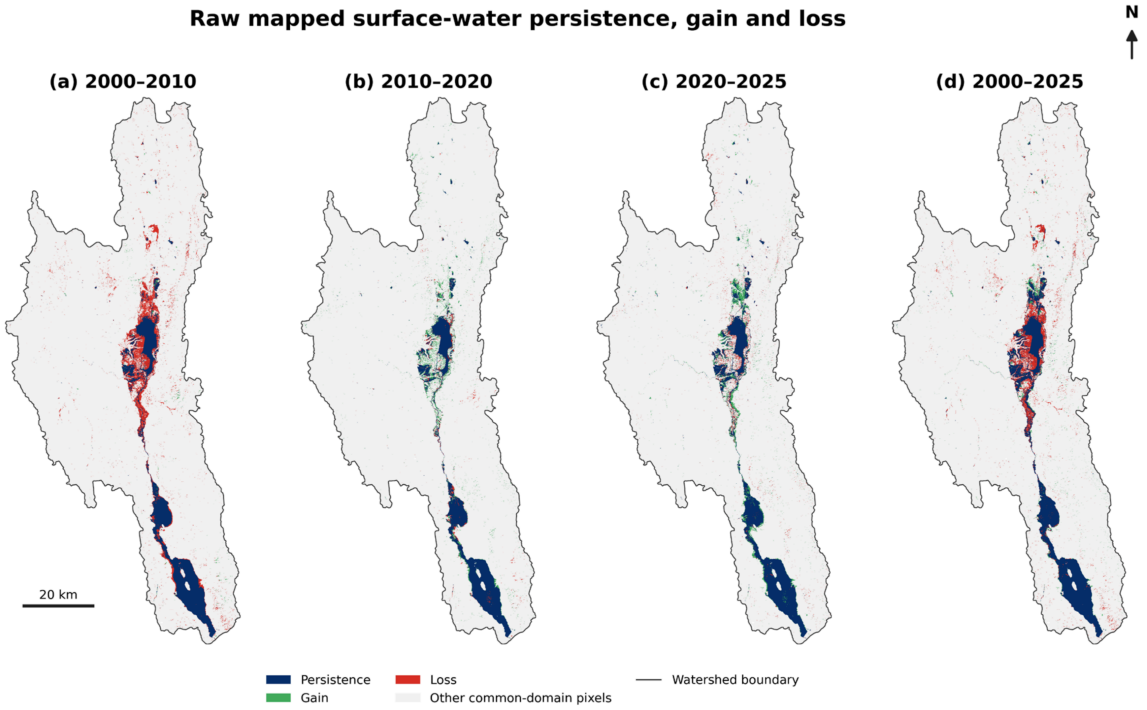

Figure 7. Raw mapped surface water persistence, gain and loss for 2000–2010, 2010–2020, 2020–2025 and 2000–2025 within the fixed common valid domain of 6,483,344 pixels (5438.20 km^2). Persistence indicates Water at both endpoints, gain indicates non-Water to Water, and loss indicates Water to non-Water. The palest grey indicates other common-domain pixels; pixels outside the domain are blank. The outline shows the rasterized watershed boundary. Maps use WGS84 geographic coordinates (EPSG:4326) with north-up orientation. All panels share the same extent and display scale; the common legend, 20 km scale bar and north arrow apply to every panel. Endpoint persistence does not establish uninterrupted water cover between the two years.

The reliability of mapped Water was assessed using the probability-sampled reference panel for 2000, 2010, 2020 and 2025. Area-weighted Water user's accuracy was 50.3%, 75.6%, 82.4% and 72.5%, respectively. These results indicate substantial commission error, particularly in 2000, and show that classification errors cannot be assumed constant across the four reference years. Consequently, differences between water masks may reflect both

surface water change and classification error. The reference-based net change estimates and sampling intervals provide a separate assessment of change, conditional on the validity of the reference labels.

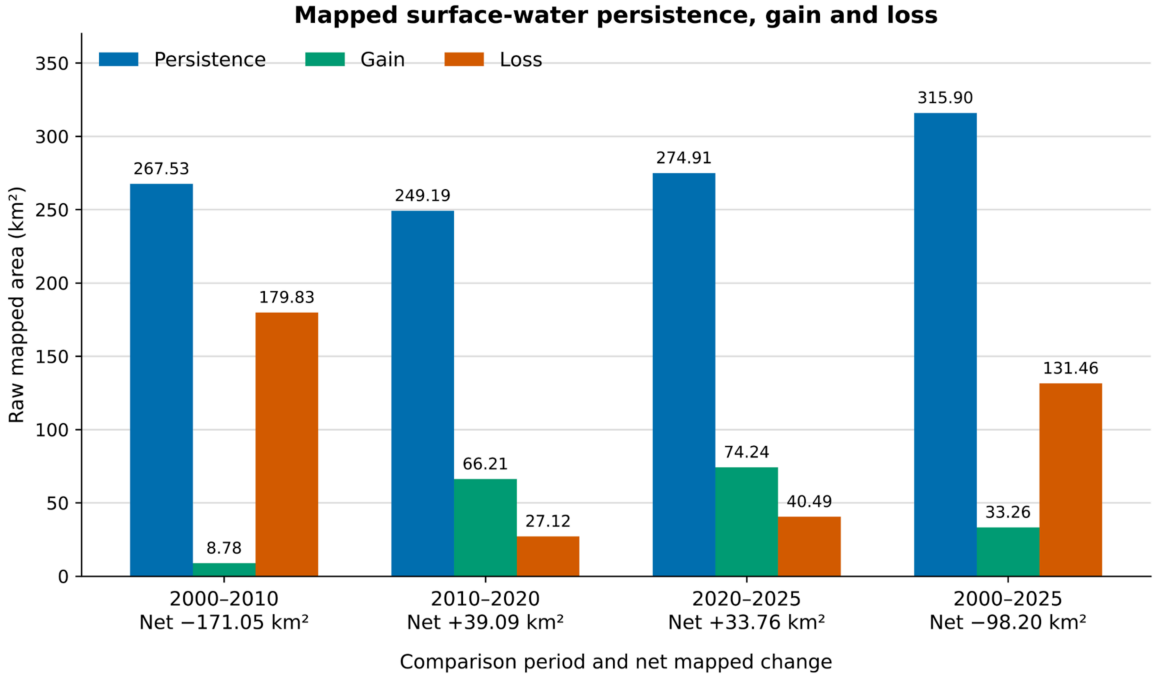

Figure 8. Raw mapped surface water persistence, gain and loss for four comparison periods within the fixed common domain of 5438.20 km^2. Gain and loss are shown as positive areas; net mapped change equals gain minus loss.

The raw water mask comparison for 2000–2010 showed a net mapped loss of 171.05 km^2, comprising 267.53 km^2 of persistent water, 8.78 km^2 of gain and 179.83 km^2 of loss. During 2010–2020, mapped water increased by 39.09 km^2, with 249.19 km^2 persisting, 66.21 km^2 gained and 27.12 km^2 lost. The 2020–2025 comparison showed a further mapped increase of 33.76 km^2, comprising 274.91 km^2 of persistence, 74.24 km^2 of gain and 40.49 km^2 of loss. Over 2000–2025, 315.90 km^2 persisted, 33.26 km^2 was gained and 131.46 km^2 was lost, yielding a net mapped decrease of 98.20 km^2. The three sub-period net changes reconcile with the full-period endpoint difference. These raw mapped increases do not establish a reference-supported recovery of surface water.

For all four comparisons, persistent water plus loss equaled the mapped water area at the start of the interval, while persistent water plus gain equaled the area at its end, to numerical precision. The 2000 baseline was consistently 447.36 km^2 in both comparisons beginning in 2000, and the 2025 endpoint was consistently 349.15 km^2 in both comparisons ending in 2025. The transition summaries therefore reconcile with the annual mapped area series over the same fixed common domain. This accounting consistency does not establish the accuracy of individual mapped transitions or remove uncertainty arising from classification and reference interpretation.

3.6. Seasonal Compositing Sensitivity

The fixed domain shared by all nine year composite combinations comprised 3,986,522 pixels (3342.52 km^2), representing 61.46% of the full common domain. Wet season predictor coverage was incomplete: missing areas accounted for 18.43%, 6.26% and 27.46% of the rasterized watershed in 2000, 2012 and 2025, respectively. Seasonal

comparisons therefore describe a restricted area with valid predictors in every assessed combination. Dry and wet season reflectance composites are shown in Figure A3.

Within this matched domain, annual composite mapped Water area was 405.84 km^2 in 2000, 299.02 km^2 in 2012 and 321.64 km^2 in 2025. The corresponding dry season areas were 413.23, 295.65 and 317.00 km^2, while wet season areas were 531.06, 324.20 and 328.76 km^2. The mapped 2000–2025 differences were −84.19 km^2 for annual composites, −96.22 km^2 for dry season composites and −202.30 km^2 for wet season composites. Thus, all three configurations produced negative endpoint differences on the matched domain, but their magnitudes depended substantially on compositing season. The mapped Water areas for the three composite configurations are compared in Figure A4.

For the wet season composites, 14.19% of matched domain pixels in 2000 and 22.35% in 2025 had at least one predictor outside the corresponding annual training sample minimum–maximum range. These descriptive diagnostics indicate differences in predictor distributions, rather than measured classification error. Without independent seasonal reference labels, seasonal surface water variation and classifier transfer error cannot be separated. These mapped sensitivity results therefore do not resolve the uncertainty in watershed-wide Water change identified by the probability-based assessment.

3.7. Lake, Reservoir and Other Water

In 2000, the mapped Water class (447.36 km^2) comprised 45.20 km^2 in the lake zone (10.1%), 191.78 km^2 in the wetland zone (42.9%), 115.00 km^2 in the reservoir zone (25.7%) and 95.38 km^2 in the other zone (21.3%) (Table 12). Within the fixed lake outline, mapped water declined slowly but monotonically from 45.20 to 42.90 km^2 (Sen's slope −0.07 km^2 yr^{-1}, 95% CI −0.10 to −0.04; Mann–Kendall $p < 0.001$; pre-whitened $p = 0.0003$); the decline continued within the Landsat 8-only period (−0.10 km^2 yr^{-1}; $p = 0.059$). Reservoir water showed no trend (115.00 and 114.79 km^2 in 2000 and 2025; $p = 0.51$). Wetland water decreased from 191.78 to 118.80 km^2 and other water from 95.38 to 72.66 km^2, but neither series had a significant monotonic trend ($p = 0.071$ and 0.48; pre-whitened $p = 0.44$ and 0.87). The wetland zone accounted for 74% of the 2000–2025 mapped decline and 62% of the 2000–2010 decline, the lake zone for only 2% and 1%. Most of this difference arises from the 2000 value: wetland water fell by 65.50 km^2 between 2000 and 2001 alone, and the 2000 value (191.78 km^2) exceeds every later year (2001–2025 maximum 159.23 km^2). Only 10 probability sample pixels fall inside the lake polygon, too few for reliable zone-level reference estimates; the reference-based change in wetland zone Water between 2000 and 2025 (+17.5 km^2, SE 17.9) does not support the mapped wetland decline.

Table 12. Mapped Water area by zone within the fixed common domain, with Mann–Kendall trend statistics for 2000–2025.

Zone	2000 (km^2)	2025 (km^2)	Change (km^2)	Sen's Slope, 95% CI (km^2 yr^{-1})	p (MK/TFPW)
Inle lake (open-water outline)	45.20	42.90	−2.29	−0.07 (−0.10 to −0.04)	<0.001/0.0003
Sanctuary wetland	191.78	118.80	−72.98	−1.02 (−2.39 to +0.12)	0.071/0.44
Moebye reservoir	115.00	114.79	−0.21	−0.09 (−0.37 to +0.19)	0.51/0.41
Outside sanctuary	95.38	72.66	−22.71	−0.28 (−1.13 to +0.50)	0.48/0.87
Total	447.36	349.15	−98.20	−1.65 (−3.88 to +0.17)	0.071/0.36

Note: Values are unadjusted mapped areas. Zones are fixed for all years (Section 2.9). Lake = Inle lake open-water polygon (45.40 km^2 of the common domain); sanctuary wetland and Moebye reservoir = Inle Wetland Bird Sanctuary north and south of 20.02°N, excluding the lake polygon. Change = 2025 minus 2000. MK = Mann–Kendall; TFPW = lag-1 trend-free pre-whitening; slope intervals are nominal. Zone values may not sum exactly to the total because of rounding. The wetland and total series have the same S (−83) because their pairwise year-to-year differences have the same sign in 299 of 325 year pairs and the net counts coincide; their slopes and pre-whitened p-values differ.

4. Discussion

4.1. Multi-Year Classifier Performance and Reliability

SVM was selected for production mapping because it achieved the highest mean Development-only grouped cross-validation macro-F1 (75.31%) compared with RF (72.33%) and XGBoost (72.30%). Outer Test scores were excluded from this selection. SVM subsequently achieved the highest mean annual internal Test OA (71.49%) and macro-F1 (71.79%). Paired spatial cluster bootstrap comparisons showed positive mean differences between SVM and both tree-based classifiers, with marginal 95% intervals excluding zero for OA and macro-F1. These results support the selection of SVM for the evaluated workflow, conditional on the fitted models and sampled Test groups. They do not establish its superiority across unsampled locations, other predictor configurations or other study regions.

Despite SVM's higher mean internal performance, classification errors remained substantial. Mean annual internal Test F1-scores were 85.79% for Forest, 75.19% for Water, 68.67% for Built-up and 57.53% for Agriculture. The separate probability-based production map assessment identified particularly low Built-up user accuracy (12.4–22.9%), indicating substantial commission error across the four reference years. Internal Test summaries and probability-based map estimates therefore need to be reported separately, with their respective sampling and uncertainty limitations. Class-specific errors are especially important when interpreting mapped area trajectories: higher overall accuracy alone does not establish the reliability of a particular class or its estimated change [28,30].

The final workflow therefore benefits from three safeguards, namely, year-specific reference samples, paired multi-model evaluation, and uncertainty reporting. Year-specific samples reduce the risk that a training set from one landscape state is transferred uncritically to another year. Paired evaluation ensures that model comparisons are made based on equivalent annual data, and spatial grouping reduces optimistic accuracy estimates caused by spatially autocorrelated samples nearby. The preceding design choices are more appropriate for a 26-year monitoring product than the maximization of the performance of any single annual map.

4.2. Long-Term Landscape Change in the Inle Lake Watershed

The long-term increase in Agriculture and decline in Forest were consistent in direction between the raw mapped totals and the probability-based reference estimates. However, changes in class area totals do not establish direct Forest-to-Agriculture conversion. During 2015–2020, raw mapped Agriculture increased by 534.08 km^2 while Forest decreased by 244.20 km^2; these endpoint differences do not imply steady annual changes. Furthermore, the Agriculture class combines terrestrial agricultural fields and floating gardens. The present classification therefore cannot distinguish their respective contributions to agricultural change or attribute mapped surface water changes specifically to floating garden expansion.

At the catchment scale, Peters et al. [6] reported that mixed agriculture increased from 37.29% to 50.36% of their catchment between 1990 and 2020, while Forest decreased from 30.40% to 16.95%. These represent changes of +13.07 and −13.45 percentage points, respectively. In our study, the reference-based Agriculture increase of 450.1 km^2 and Forest decrease of 631.5 km^2 during 2000–2025 correspond to +8.28 and −11.61 percentage points of the fixed 5438.20 km^2 domain. The studies therefore agree in the direction of agricultural expansion and forest loss. Their magnitudes are not directly equivalent, however: Peters et al. delineated a 4368.25 km^2 catchment and distinguished eight classes, including separate mixed agriculture, rice, floating garden and range brush classes, whereas our four-class scheme aggregates several of these landscape components. Dividing these endpoint changes by the elapsed periods gives average changes in catchment share of

approximately +0.44 percentage points yr^{-1} for mixed agriculture and −0.45 percentage points yr^{-1} for Forest in Peters et al. [6], compared with +0.33 and −0.46 percentage points yr^{-1}, respectively, in our study; these descriptive averages do not imply constant annual change or represent fitted trend slopes. Differences in boundaries, observation periods, class definitions and accuracy assessment prevent this directional agreement from serving as independent validation of our change estimates.

During 2020–2025, raw mapped Agriculture decreased by 132.75 km^2 and Forest decreased by 103.83 km^2. The maps therefore do not indicate forest recovery during this interval. These mapped differences alone cannot distinguish persistent land cover change from crop and fallow dynamics, phenological variation or classification error. Short-period interpretations should consequently consider the paired reference-based change estimates and their sampling intervals, while acknowledging that those intervals exclude reference interpretation error. A decrease in mapped agricultural area should not, by itself, be interpreted as forest regeneration or ecological recovery.

The raw mapped Built-up decline of 321.19 km^2 contrasted with the probability-based reference estimate of an increase of 139.2 km^2 (95% sampling interval: 23.5–254.9 km^2) during 2000–2025. Built-up user's accuracy ranged from 12.4% to 22.9% across the four reference years, indicating substantial commission errors. The raw mapped decline should therefore not be interpreted as settlement contraction. The reference-based increase remains conditional on reference label validity, and its sampling interval excludes uncertainty from reference interpretation.

The Landsat 8-only analysis provides an important qualification to the multi-decadal interpretation. During 2013–2025, mapped Forest retained a negative trend (Sen's slope: -24.89 km^2 yr^{-1}; trend-free pre-whitening $p = 0.0020$), whereas Agriculture showed no significant monotonic trend (Sen's slope: -1.21 km^2 yr^{-1}; pre-whitening $p = 0.6312$), despite a positive endpoint difference of 353.81 km^2. Thus, agricultural expansion between assessed endpoints should not be interpreted as a persistent annual increase throughout either the OLI period or the complete record. The within-OLI forest decline shows that a negative mapped trajectory also occurs without crossing the legacy-to-OLI sensor boundary. However, restricting the period does not quantify sensor-induced bias in the full record or eliminate effects of observation availability and classification error. The reference-based endpoint changes and the raw annual trajectories therefore provide distinct evidence and should be interpreted separately.

The reference-based estimates add further qualifications in timing. Almost all of the reference-supported agricultural increase and forest loss occurred between 2000 and 2010, both Landsat 5 TM reference years, whereas changes during 2010–2020 and 2020–2025 were not significant (Table A4). The supported land cover change therefore does not coincide with the TM-to-OLI transition, and the large mapped redistribution during 2015–2020 (Agriculture +534.08 km^2, Forest −244.20 km^2) is not supported by the reference estimates for 2010–2020. Because 2000 labels were more often medium confidence than later labels, the 2000–2010 estimates are also the most exposed to interpretation error.

4.3. Surface Water Variability and Hydrological Interpretation

Raw watershed-wide mapped surface water area decreased by 98.20 km^2 between 2000 and 2025, but the annual series did not show a statistically significant monotonic trend at the nominal 0.05 level (Mann–Kendall $\tau = -0.255$, $p = 0.0707$; lag-1 trend-free pre-whitening $p = 0.3624$). The probability-based reference assessment estimated a full-period Water change of +42.3 km^2, with a 95% sampling interval of −23.5 to +108.0 km^2. Thus, this assessment established neither a clear full-period increase nor a decrease. The difference between the raw mapped and reference-based endpoint changes highlights

the influence of classification error on the inferred direction and magnitude of surface water change. In particular, the mapped 2000 Water area (447.36 km^2) is the largest in the series and coincides with the lowest Water user's accuracy (50.3%), which indicates that commission error inflates the 2000 map and contributes to the raw mapped decline. The zone decomposition (Section 3.7) supports this: 74% of the mapped 2000–2025 decline lies in the sanctuary wetlands, largely as a 2000-to-2001 drop, whereas the Moebye reservoir shows no trend, so reservoir operations do not explain the mapped decline. Hydrological interpretation therefore requires consideration of classification and reference uncertainty alongside the mapped temporal variability.

Earlier studies provide quantitative context but differ in spatial extent, observation period and class definitions. Sidle et al. [7] reported that Inle lake 's open-water area decreased from 69.10 km^2 in 1935 to 46.69 km^2 in 2000, a reduction of 32.4%. Peters et al. [6] reported a 21.33% reduction in lake open-water area between 1990 and 2020 and estimated approximately 38 km^2 of open water in 2020. These lake-specific values are not comparable with our watershed-wide Water totals, which include Moebye reservoir, channels, ponds and other water bodies, but they are comparable with our lake zone series (Section 3.7). Mapped water within the fixed open-water outline was 45.20 km^2 in 2000 and 43.14 km^2 in 2020, compared with 46.69 km^2 (Sidle et al., 2007 [7]) and about 38 km^2 (Peters et al., 2023 [6]). The 2000 values agree within 1.5 km^2. Our 2020 value is about 5 km^2 larger, probably because a 30 m annual median classification labels mixed open-water and floating vegetation pixels as Water, and because the two studies delineated the lake differently. Both records nevertheless show open-water loss inside the lake: our series declines monotonically (-0.07 km^2 yr^{-1}; $p < 0.001$), equivalent to 2.29 km^2 (5.1%) between 2000 and 2025. Because the outline is fixed, this value measures encroachment of non-water cover within the 2000 open-water area and excludes any loss outside it.

Floating garden estimates also require separate interpretation. Michalon et al. [5] reported approximately 35 km^2 in 2002 and 33 km^2 in 2014, while Peters et al. [6] estimated approximately 32 km^2 in 2020. Our Agriculture class combines cultivated floating gardens with terrestrial agricultural and fallow fields. Consequently, the reference-based agricultural increase of 450.1 km^2 during 2000–2025 cannot be interpreted as floating garden expansion. Assessing whether lake open-water loss coincides with floating garden change would require separate delineation of these components using comparable spatial boundaries, class definitions and observation periods.

The raw water mask comparisons showed mapped increases of 39.09 km^2 during 2010–2020 and 33.76 km^2 during 2020–2025. However, the reference-based estimate for 2020–2025 was -4.6 km^2, with a 95% sampling interval of -43.9 to $+34.8$ km^2 and therefore did not establish a recovery. Mapped gains should consequently be interpreted as changes in classified surface water extent rather than confirmed hydrological improvement. Without corresponding observations of water level, discharge, rainfall and reservoir operations, these mapped changes cannot be attributed to a particular climatic or management driver.

The seasonal comparison further demonstrates that mapped Water extent depends on the temporal compositing window (Section 3.6; Figures A3 and A4). Annual median reflectance summarizes available valid observations; the resulting classified Water area is not an estimate of annual mean inundation or seasonal maximum or minimum extent. The larger wet season mapped areas may reflect actual surface water and vegetation variation, uneven observation availability, and transfer of annual trained classifiers to seasonal predictors. Because the matched comparison covers only 61.46% of the full common domain and lacks independent seasonal reference labels, these contributions cannot be separated. The experiment therefore demonstrates compositing sensitivity without establishing the magnitude of actual seasonal inundation change.

4.4. Uncertainty and Limitations

Annual Landsat median composites remain sensitive to residual cloud and shadow contamination, mixed pixels, floating vegetation, turbid water, wet soil and phenological variation. They summarize available observations across the calendar year and do not represent maximum or minimum inundation, a fixed hydrological season, or water level. Differences among sensors and the seasonal distribution of valid observations may also influence the annual classifications; their contributions have not been isolated from land-cover change in this analysis. Area-weighted accuracy was lower in the Landsat 5 TM reference years (58.7% in 2000 and 64.3% in 2010) than in the Landsat 8 OLI years (71.4% in 2020 and 69.8% in 2025), which is consistent with map quality that depends on sensor or epoch. The reference-based estimates are less affected, because they rely on interpreted labels and design weights: map error, including sensor-related error, reduces their precision but does not bias them. Because labels were assigned first for 2025 and then checked backwards year by year, and because high-resolution imagery did not cover the whole study area for most of 2000–2010, some changes may be under-recorded, and some early labels were confirmed from adjacent-year imagery rather than observed directly (Section 2.4). This affects the 2000–2010 reference estimates most. Historical reference interpretation introduces additional uncertainty, particularly where imagery is limited or class boundaries are ambiguous. The seasonal sensitivity analysis retained 61.46% of the full common domain; the remaining 38.54% was excluded through the shared validity requirement and the documented control exclusions. This retained area is not a probability sample of the original domain, so its seasonal results cannot be extrapolated to excluded locations. Sparse and spatially uneven wet season observations also limit the representativeness of the seasonal medians. Furthermore, applying annual-trained SVM models without independent seasonal validation leaves seasonal classification error unquantified. The seasonal comparison therefore provides a conditional assessment of compositing sensitivity rather than validated seasonal water area estimates.

Although reference interpretation was divided among the three authors by year and the resulting labels were cross-checked within the team, this procedure did not provide separately recorded repeat interpretations of the same samples. Only the final reference labels were retained. Inter-operator agreement was therefore not measured. The confidence levels recorded at interpretation show where labels were least certain (31% of 2000 labels were medium confidence, compared with 8% in 2010), and the low-confidence sensitivity analysis (Section 3.4) bounds only the effect of labels the interpreters themselves flagged; it does not capture errors in High- or Medium-confidence labels. The reported sampling intervals are conditional on the final reference labels and do not incorporate uncertainty arising from interpretation errors.

The Water class includes Moebye reservoir, channels, ponds and other water bodies within the mapped watershed domain, alongside Inle lake. Lake, wetland, reservoir and other water were separated (Section 3.7), but the zones use a fixed lake outline and a sanctuary boundary rather than year-specific delineations, the wetland zone was not divided further, and changes in each zone are not attributed to monsoon variability or reservoir operations. Similarly, the Agriculture class combines terrestrial fields and cultivated floating gardens. Changes in this aggregate class do not establish the contribution of floating garden expansion to lake surface change.

Annual mapped area and water transition comparisons were restricted to a fixed common valid domain of 6,483,344 pixels (5438.20 km^2). These estimates therefore describe that domain rather than the complete watershed boundary, including any excluded pixels. The probability-based reference assessment covered four years using the same 600 sampled pixels; the resulting 2400 pixel year records do not represent 2400 independent locations or

independently validate all 26 annual maps. Reported sampling intervals remain conditional on reference label validity and exclude interpretation error. Furthermore, the interpolated annual correction patterns are sensitivity scenarios rather than observed annual classification errors. Their trend test results do not incorporate uncertainty in the estimated corrections.

Finally, the analysis is descriptive instead of causal. No attempt is made in this research to infer that a specific climate variable or land cover class directly caused an annual change in water area. To establish the causal mechanism behind such a change requires additional hydrometeorological observations, lagged climate analyses, water level records, runoff or inflow data, and potential hydrological modeling. Instead, the strength of the present study lies in a consistent annual record of remote sensing with explicit model comparison and uncertainty characterization.

5. Conclusions

This study constructed annual LULC and watershed-wide surface water conditions in the Inle lake watershed for 2000–2025 using Landsat annual composites and three machine learning classifiers. SVM was selected for production mapping using Development-only grouped cross-validation performance. Its 26-year mean internal Test OA, kappa and macro-F1 were 71.49%, 0.620 and 71.79%, respectively. Paired spatial cluster bootstrap comparisons indicated positive conditional mean OA and macro-F1 differences between SVM and both tree-based classifiers. These internal evaluations were complemented by a separate probability-based assessment of production map accuracy, class areas and net changes for four reference years.

Within the fixed common domain, the reference-based assessment estimated an agricultural increase of 450.1 km^2 (95% sampling interval: +203.4 to +696.8 km^2) and a forest decrease of 631.5 km^2 (−857.3 to −405.7 km^2) between 2000 and 2025. These estimates support agricultural expansion and forest loss. Water change was estimated at +42.3 km^2 (−23.5 to +108.0 km^2), establishing neither a clear full-period increase nor a decrease. Within the fixed open water outline of Inle lake, mapped water declined slowly but consistently (−0.07 km^2 yr^{-1}; $p < 0.001$), the Moebye reservoir showed no trend, and most of the mapped watershed decline occurred in the surrounding sanctuary wetlands. The reference assessment also did not establish a recovery during 2020–2025. Built-up user accuracy remained low across the four reference years (12.4–22.9%); its raw mapped decline therefore should not be interpreted as confirmed settlement contraction. These findings demonstrate the importance of distinguishing raw mapped trajectories from reference-based change estimates and their sampling uncertainty.

The annual mapped record adds temporal detail between the four reference years, while annual fluctuations and individual transitions still require further validation. The transferable methodological lesson is to interpret mapped patterns alongside probability-based estimates of change, with their spatial support and uncertainty stated explicitly. This distinction identifies changes supported by reference observations and separates them from patterns requiring further evidence. A consistent annual map series alone does not establish classification accuracy or the causes of environmental change.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/rs18213837/s1, Data S1: Inventory of Landsat Collection 2 Level 2 scenes used for annual compositing, Inle Lake watershed, Myanmar, 2000–2025.

Author Contributions: Conceptualization, L.M.K.K.; methodology, L.M.K.K.; software, L.M.K.K. and P.T.H.; validation, L.M.K.K., W.S.M. and P.T.H.; data curation, L.M.K.K.; writing—original draft preparation, L.M.K.K.; writing—review and editing, L.M.K.K., W.S.M. and P.T.H.; supervision, W.S.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The study used publicly available Landsat Collection 2 Level 2 surface reflectance data accessed through Google Earth Engine. The reference labels (purposive pool and probability panel), the Landsat scene inventory (Data S1), the analysis code and the derived annual LULC maps and surface water transition products are available from the corresponding author on reasonable request. Reproducible analysis scripts were implemented in Google Earth Engine and Python/Google Colab.

Acknowledgments: The authors thank the developers and data providers of the Landsat archive, Google Earth Engine, scikit-learn, and XGBoost, and acknowledge the contributions of previous environmental and socio-hydrological research conducted in the Inle lake region. The authors express our gratitude to the anonymous reviewers for their careful reading of our manuscript and their insightful comments and suggestions for further improvement.

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

Appendix A. Reference Sample Quality Control and Sample Counts

Appendix A.1. Quality-Control Rules for Reference Samples

Input validation failures halt processing for correction and are not counted as raster exclusions. Raster decisions apply to the predictor pixel containing each point.

Check	Decision rule or threshold	Outcome in this run
Annual file and fields	One CSV per year; exactly 1000 complete records; required fields present.	Passed input validation
Identity and geometry	SOURCE_POINT_ID 1-1000 once; unique POINT_ID and coordinates; complete and consistent Pair_ID and POINT_TYPE metadata.	Passed input validation
Coordinates and classes	Finite latitude −90 to 90 and longitude −180 to 180; exact recognized class names.	Passed input validation
Cross year identity	Same ID set, source IDs and point types; coordinate equality with absolute tolerance 1×10^{-10} degrees and relative tolerance 0.	Passed; tolerance checks file consistency, not positional accuracy
Raster extent	Reject if row < 0 or row $\geq$ raster height, or column < 0 or column $\geq$ raster width.	0 excluded
Mask or NoData	Reject if any predictor band is masked or flagged NoData by raster masking.	26 excluded
Nonfinite predictors	Reject if any unmasked predictor is NaN, +Inf or −Inf. All 12 predictors must be finite.	0 excluded
Conflicting pixel labels	If raster-valid records occupy the same row/column with different labels, reject all conflicting records.	0 excluded
Redundant pixel records	For same-pixel, same-label records, retain one using stored point-type priority, then lowest source ID. The priority ordering is recorded in the QC script.	26 excluded
Class sample safeguard	At least 50 valid records in each class/year before splitting; otherwise halt. This is not a purity threshold or a per-fold requirement.	Passed in all years
Spectral plausibility	No reflectance range, index range, vegetation cover or image purity numerical cutoff.	No such exclusion rule
Visual interpretation	Clearly identifiable classes preferred; mixed or unclear boundaries avoided as far as practicable.	Qualitative selection; not counted as a measured QC error rate

Appendix A.2. Annual Reference-Sample Counts

Each year started with 1000 candidate records. Blocks and groups below count those represented by valid records after QC. D = Development; T = Test. Five grouped folds were used every year; fallback count was zero.

Year	Valid	Excluded	D	T	Blocks	Groups	Folds	Fallback
2000	998	2	700	298	449	381	5	No
2001	998	2	700	298	449	381	5	No
2002	998	2	700	298	449	381	5	No
2003	998	2	700	298	449	381	5	No
2004	998	2	700	298	449	381	5	No
2005	998	2	700	298	449	381	5	No
2006	998	2	700	298	449	381	5	No
2007	998	2	700	298	449	381	5	No
2008	998	2	700	298	449	381	5	No
2009	998	2	700	298	449	381	5	No
2010	998	2	700	298	449	381	5	No
2011	998	2	700	298	449	381	5	No
2012	998	2	700	298	449	381	5	No
2013	998	2	700	298	449	381	5	No
2014	998	2	700	298	449	381	5	No
2015	998	2	700	298	449	381	5	No
2016	998	2	700	298	449	381	5	No
2017	998	2	700	298	449	381	5	No
2018	998	2	700	298	449	381	5	No
2019	998	2	700	298	449	381	5	No
2020	998	2	700	298	449	381	5	No
2021	998	2	700	298	449	381	5	No
2022	998	2	700	298	449	381	5	No
2023	998	2	700	298	449	381	5	No
2024	998	2	700	298	449	381	5	No
2025	998	2	700	298	449	381	5	No
Total	25,948	52	18200	7748	—	—	—	0 years

Note: Each year began with 1000 candidate records. D = Development; T = Test. Block and group counts refer to each annual dataset. Five-fold grouped cross-validation was used every year; no non-grouped fallback was applied.

Appendix A.3. Class-Specific Reference-Sample Counts

Within each class cell, values are C/V/D/T: Candidate/Valid/Development/Test records. The order is identical in every cell. V = D + T. These are annual reference sample counts, not mapped areas or watershed class proportions.

Year	Water C/V/D/T	Built Up C/V/D/T	Agriculture C/V/D/T	Forest C/V/D/T
2000	246/246/170/76	246/245/173/72	256/256/178/78	252/251/179/72
2001	246/246/170/76	248/247/175/72	256/256/178/78	250/249/177/72
2002	222/222/156/66	250/249/173/76	276/276/192/84	252/251/179/72
2003	222/222/156/66	250/249/173/76	276/276/192/84	252/251/179/72
2004	222/222/156/66	250/249/173/76	276/276/192/84	252/251/179/72
2005	244/244/168/76	250/249/175/74	256/256/182/74	250/249/175/74
2006	230/230/160/70	250/249/175/74	270/270/190/80	250/249/175/74
2007	230/230/160/70	250/249/175/74	270/270/190/80	250/249/175/74
2008	224/224/154/70	250/249/175/74	276/276/196/80	250/249/175/74
2009	224/224/154/70	250/249/175/74	276/276/196/80	250/249/175/74
2010	230/230/160/70	250/249/175/74	270/270/188/82	250/249/177/72
2011	230/230/160/70	250/249/175/74	270/270/188/82	250/249/177/72
2012	224/224/158/66	254/253/177/76	272/272/188/84	250/249/177/72
2013	222/222/156/66	254/253/177/76	274/274/190/84	250/249/177/72
2014	218/218/156/62	254/253/177/76	278/278/190/88	250/249/177/72
2015	246/246/170/76	250/249/175/74	254/254/178/76	250/249/177/72
2016	242/242/168/74	250/249/175/74	258/258/180/78	250/249/177/72
2017	242/242/166/76	250/249/175/74	258/258/182/76	250/249/177/72
2018	244/244/168/76	250/249/175/74	256/256/180/76	250/249/177/72
2019	242/242/166/76	250/249/175/74	258/258/182/76	250/249/177/72
2020	240/240/164/76	250/249/175/74	260/260/184/76	250/249/177/72
2021	234/234/160/74	250/249/175/74	266/266/188/78	250/249/177/72
2022	234/234/160/74	250/249/175/74	266/266/188/78	250/249/177/72
2023	236/236/162/74	250/249/175/74	264/264/186/78	250/249/177/72
2024	236/236/162/74	250/249/175/74	264/264/186/78	250/249/177/72
2025	250/250/172/78	250/249/175/74	250/250/176/74	250/249/177/72
Total	6080/6080/4212/1868	6506/6480/4548/1932	6906/6906/4840/2066	6508/6482/4600/1882

Appendix B. Additional Classification and Compositing Diagnostics

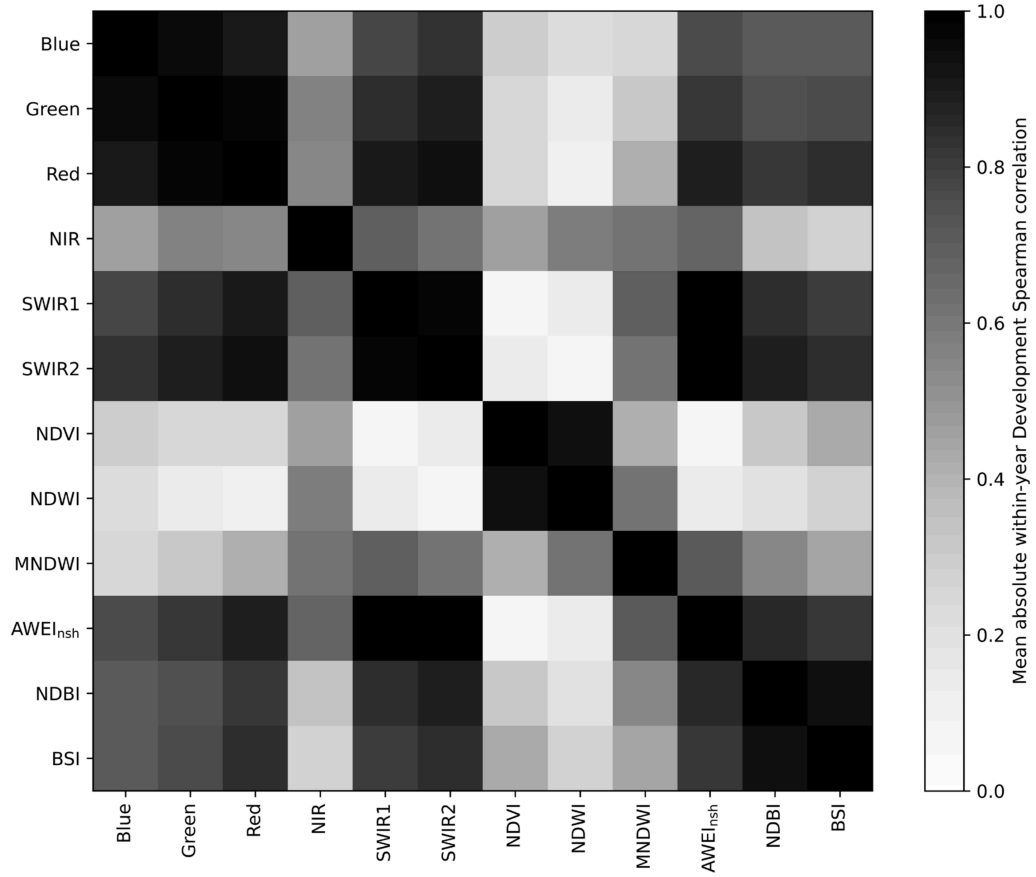

Figure A1. Mean absolute Spearman correlations among the twelve predictors. Correlations were calculated separately from each year's Development samples and their absolute values averaged across 2000–2025. Darker shading indicates stronger absolute association. The matrix summarizes predictor redundancy within the Development data.

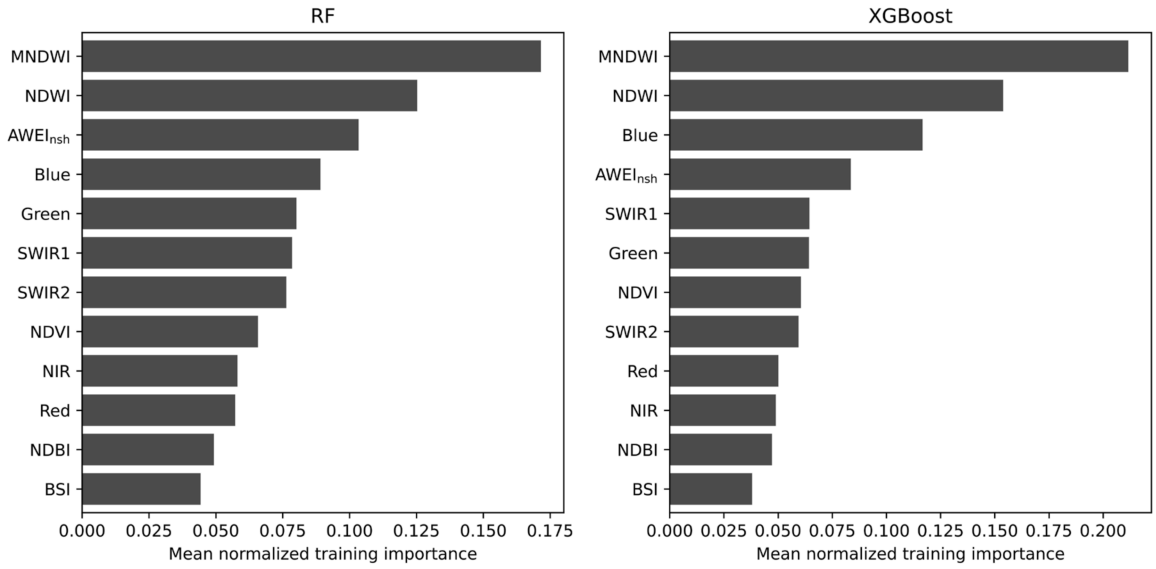

Figure A2. Mean predictor importance across the 26 annual twelve-predictor models fitted to Development samples. RF importance represents mean decrease in impurity; XGBoost importance represents normalized total gain. Bars show unweighted means of normalized importance scores. The two importance measures are model-specific and describe the fitted models rather than causal effects.

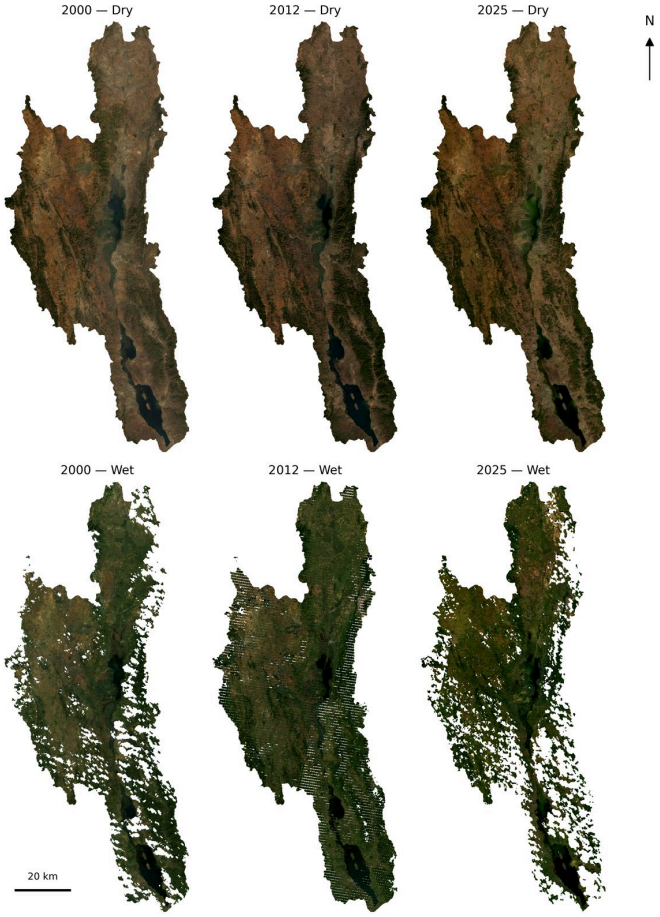

Figure A3. Dry and wet season Landsat median reflectance composites for 2000, 2012 and 2025, displayed in true color using red, green and blue bands with a common reflectance stretch. Dry season composites combine January–April and November–December observations within each calendar year; wet season composites use May–October observations. White areas indicate missing RGB data or locations outside the watershed. Panels show the full watershed extent; quantitative seasonal comparisons use the smaller fixed matched domain of 3342.52 km^2.

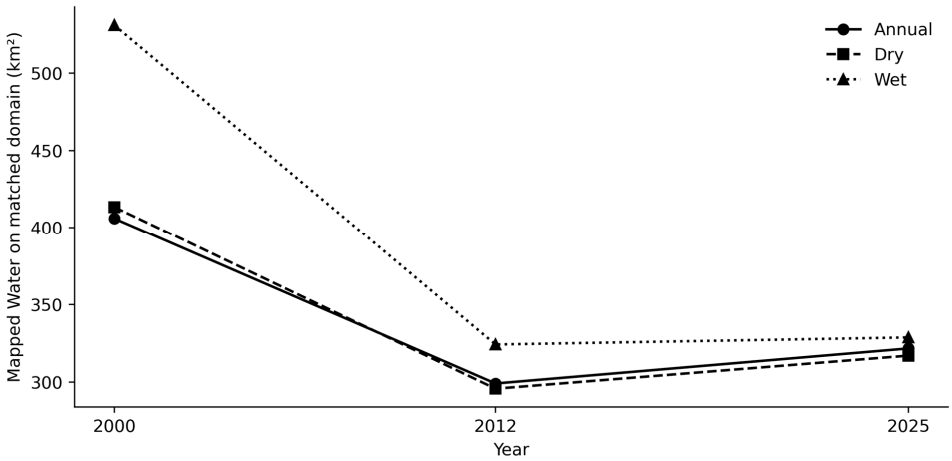

Figure A4. Mapped Water area from annual, dry season and wet season composites in 2000, 2012 and 2025 within the fixed matched domain of 3342.52 km^2. Each year's annual production SVM was held fixed across composite configurations. Points represent unadjusted classified map areas; connecting lines link the assessed years and do not represent annual observations or fitted trends. Seasonal classification accuracy was not independently evaluated.

Appendix C. Landsat Scene Availability

Table A1. Inventory of Landsat scenes meeting the watershed, calendar year and scene-level cloud cover (<30%) filters, summarized by sensor and season.

Year	Total Scenes	L5	L7	L8	Dry Scenes	Wet Scenes	Unique Dates
2000	36	36	0	0	30	6	20
2001	21	21	0	0	18	3	12
2002	38	0	38	0	30	8	24
2003	42	6	36	0	37	5	22
2004	41	41	0	0	28	13	22
2005	39	39	0	0	28	11	23
2006	30	30	0	0	26	4	17
2007	29	29	0	0	26	3	15
2008	33	33	0	0	32	1	17
2009	40	40	0	0	34	6	20
2010	23	23	0	0	20	3	13
2011	14	14	0	0	13	1	8
2012	39	0	39	0	28	11	23
2013	23	0	0	23	17	6	14
2014	46	0	0	46	37	9	26
2015	38	0	0	38	34	4	20
2016	46	0	0	46	36	10	26
2017	43	0	0	43	36	7	24
2018	48	0	0	48	38	10	28
2019	59	0	0	59	43	16	31
2020	49	0	0	49	39	10	26
2021	41	0	0	41	36	5	22
2022	39	0	0	39	37	2	21
2023	42	0	0	42	34	8	22
2024	48	0	0	48	44	4	26
2025	45	0	0	45	38	7	24

Note: Dry comprises January–April and November–December within the same calendar year; Wet comprises May–October. Scene counts include separate Landsat products acquired on the same date across different path/row tiles. Unique dates count distinct acquisition dates. These scene counts do not represent valid observation counts at individual pixels. The inventory was compiled from the Landsat archive and does not independently establish the exact scene membership of every annual composite.

Appendix D. Internal SVM Confusion Matrices

Table A2. Internal SVM outer Test confusion counts for 2000, 2012 and 2025. Rows indicate reference classes and columns indicate predicted classes.

Year	Reference Class	Predicted Water	Predicted Built-Up	Predicted Agriculture	Predicted Forest
2000	Water	61	7	7	1
2000	Built-up	7	48	10	7
2000	Agriculture	15	15	35	13
2000	Forest	1	3	5	63
2012	Water	50	7	8	1
2012	Built-up	3	48	23	2
2012	Agriculture	13	15	53	3
2012	Forest	0	2	10	60
2025	Water	51	10	15	2
2025	Built-up	4	59	10	1
2025	Agriculture	5	14	51	4
2025	Forest	0	0	2	70

Note: Each year contains 298 outer Test observations. Models were fitted to Development samples, with Test observations excluded from model fitting and hyperparameter selection. These unweighted counts describe internal evaluation on purposively selected, spatially grouped samples. They are distinct from the probability-based production map assessment and do not provide population-level accuracy estimates for 2012.

Appendix E. Reference-Based Class Areas and Changes

Table A3. Reference-based class areas (km^2, ± sampling standard error) from the stratified probability sample, and unadjusted mapped areas, for the four reference years.

Class	Estimate	2000	2010	2020	2025
Water	Reference ± SE	254 ± 31	259 ± 32	301 ± 31	296 ± 29
	Mapped	447	276	315	349
Built-up	Reference ± SE	219 ± 46	316 ± 55	278 ± 49	358 ± 58
	Mapped	1016	810	492	695
Agriculture	Reference ± SE	2524 ± 113	2975 ± 107	2990 ± 101	2974 ± 103
	Mapped	1671	1989	2703	2571
Forest	Reference ± SE	2442 ± 107	1888 ± 98	1869 ± 93	1810 ± 95
	Mapped	2304	2362	1928	1824

Note: Reference = stratified estimator applied to the 600-pixel probability sample (Section 2.6); SE = sampling standard error, excluding interpretation error. Mapped = unadjusted SVM map totals within the common domain (5438.20 km^2). Medium-confidence labels made up 31% of 2000 labels, 8% of 2010, 3% of 2020 and 13% of 2025. Values rounded to whole km^2.

Table A4. Reference-based net changes (km^2, with approximate 95% sampling intervals) and unadjusted mapped changes for each sub-period and for 2000–2025.

Class	Estimate	2000–2010	2010–2020	2020–2025	2000–2025
Water	Reference (95% interval)	+5 (−66 to +76)	+42 (−19 to +103)	−5 (−44 to +35)	+42 (−24 to +108)
	Mapped	−171	+39	+34	−98
Built-up	Reference (95% interval)	+98 (−12 to +207)	−39 (−102 to +24)	+80 (+9 to +151)	+139 (+24 to +255)
	Mapped	−205	−319	+203	−321
Agriculture	Reference (95% interval)	+451 (+218 to +685)	+16 (−169 to +200)	−17 (−150 to +117)	+450 (+203 to +697)
	Mapped	+319	+714	−133	+900
Forest	Reference (95% interval)	−554 (−765 to −343)	−19 (−188 to +151)	−59 (−182 to +64)	−632 (−857 to −406)
	Mapped	+58	−435	−104	−481

Note: Paired changes at the same sample pixels; intervals are estimate ± 1.96 SE and exclude interpretation error. The 2000–2025 reference column corresponds to Table 10 (rounded to whole km^2).

References

1. Karki, S.; Thandar, A.M.; Uddin, K.; Tun, S.; Aye, W.M.; Aryal, K.; Kandel, P.; Chettri, N. Impact of Land Use Land Cover Change on Ecosystem Services: A Comparative Analysis on Observed Data and People's Perception in Inle Lake, Myanmar. *Environ. Syst. Res.* **2018**, *7*, 25. [CrossRef]
2. Downing, J.A.; Prairie, Y.T.; Cole, J.J.; Duarte, C.M.; Tranvik, L.J.; Striegl, R.G.; McDowell, W.H.; Kortelainen, P.; Caraco, N.F.; Melack, J.M.; et al. The Global Abundance and Size Distribution of Lakes, Ponds, and Impoundments. *Limnol. Oceanogr.* **2006**, *51*, 2388–2397. [CrossRef]
3. Foley, J.A.; DeFries, R.; Asner, G.P.; Barford, C.; Bonan, G.; Carpenter, S.R.; Chapin, F.S.; Coe, M.T.; Daily, G.C.; Gibbs, H.K.; et al. Global Consequences of Land Use. *Science* **2005**, *309*, 570–574. [CrossRef] [PubMed]
4. Vörösmarty, C.J.; McIntyre, P.B.; Gessner, M.O.; Dudgeon, D.; Prusevich, A.; Green, P.; Glidden, S.; Bunn, S.E.; Sullivan, C.A.; Liermann, C.R.; et al. Global Threats to Human Water Security and River Biodiversity. *Nature* **2010**, *467*, 555–561. [CrossRef] [PubMed]
5. Michalon, M.; Gunnell, Y.; Lejot, J.; Mialhe, F.; Aung, T. Accelerated Degradation of Lake Inle (Myanmar): A Baseline Study for Environmentalists and Developers. *Land Degrad. Dev.* **2019**, *30*, 928–941. [CrossRef]
6. Peters, K.; Wagner, P.D.; Phyo, E.W.; Zin, W.W.; Kyi, C.C.T.; Fohrer, N. Spatial and Temporal Assessment of Human-Water Interactions at the Inle Lake, Myanmar: A Socio-Hydrological DPSIR Analysis. *Environ. Monit. Assess.* **2023**, *195*, 220. [CrossRef] [PubMed]
7. Sidle, R.C.; Ziegler, A.D.; Vogler, J.B. Contemporary Changes in Open Water Surface Area of Lake Inle, Myanmar. *Sustain. Sci.* **2007**, *2*, 55–65. [CrossRef]
8. Wulder, M.A.; Masek, J.G.; Cohen, W.B.; Loveland, T.R.; Woodcock, C.E. Opening the Archive: How Free Data Has Enabled the Science and Monitoring Promise of Landsat. *Remote Sens. Environ.* **2012**, *122*, 2–10. [CrossRef]
9. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [CrossRef]

10. White, J.C.; Wulder, M.A.; Hobart, G.W.; Luther, J.E.; Hermosilla, T.; Griffiths, P.; Coops, N.C.; Hall, R.J.; Hostert, P.; Dyk, A.; et al. Pixel-Based Image Compositing for Large-Area Dense Time Series Applications and Science. *Can. J. Remote Sens.* **2014**, *40*, 192–212. [CrossRef]
11. Donchyts, G.; Baart, F.; Winsemius, H.; Gorelick, N.; Kwadijk, J.; Van De Giesen, N. Earth's Surface Water Change over the Past 30 Years. *Nat. Clim. Chang.* **2016**, *6*, 810–813. [CrossRef]
12. Pekel, J.F.; Cottam, A.; Gorelick, N.; Belward, A.S. High-Resolution Mapping of Global Surface Water and Its Long-Term Changes. *Nature* **2016**, *540*, 418–422. [CrossRef] [PubMed]
13. Belgiu, M.; Drăgu, L. Random Forest in Remote Sensing: A Review of Applications and Future Directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]
14. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13 August 2016*; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794.
15. Mountrakis, G.; Im, J.; Ogole, C. Support Vector Machines in Remote Sensing: A Review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 247–259. [CrossRef]
16. Meyer, H.; Reudenbach, C.; Wöllauer, S.; Nauss, T. Importance of Spatial Predictor Variable Selection in Machine Learning Applications -Moving from Data Reproduction to Spatial Prediction. *Ecol. Model.* **2019**, *411*, 108815. [CrossRef]
17. Michalon, M. *The Inlay Lake Region in the Mists of Tourism Governance*; Circle Oxford: Manchester, UK, 2017. [CrossRef]
18. ICIMOD; MONREC. *A Multi-Dimensional Assessment of Ecosystems and Ecosystem Services at Inle Lake, Myanmar*; ICIMOD Working Paper 2017/17; ICIMOD: Kathmandu, Nepal; MONREC: Naypyidaw, Myanmar, 2017.
19. McFeeters, S. The Use of the Normalized Difference Water Index (NDWI) in the Delineation of Open Water Features. *Int. J. Remote Sens.* **1996**, *17*, 1425–1432. [CrossRef]
20. Xu, H. Modification of Normalised Difference Water Index (NDWI) to Enhance Open Water Features in Remotely Sensed Imagery. *Int. J. Remote Sens.* **2006**, *27*, 3025–3033. [CrossRef]
21. Feyisa, G.L.; Melby, H.; Fensholt, R.; Proud, S.R. Automated Water Extraction Index: A New Technique for Surface Water Mapping Using Landsat Imagery. *Remote Sens. Environ.* **2014**, *140*, 23–35. [CrossRef]
22. Zha, Y.; Gao, J.; Ni, S. Use of Normalized Difference Built-up Index in Automatically Mapping Urban Areas from TM Imagery. *Int. J. Remote Sens.* **2003**, *24*, 583–594. [CrossRef]
23. Rikimaru, A.; Roy, P.S.; Miyatake, S. Tropical Forest Cover Density Mapping. *Trop. Ecol.* **2002**, *43*, 39–47.
24. Rouse, R.W.H.; Haas, J.A.W.; Deering, D.W. *Monitoring Vegetation Systems in the Great Plains with ERTS*; NASA Paper-A20; NASA: Washington, DC, USA, 1973.
25. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
26. Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Meas.* **1960**, *20*, 37–46. [CrossRef]
27. Congalton, R.G.; Green, K. Assessing the Accuracy of Remotely Sensed Data. In Proceedings of the Seventh Annual Forest Inventory and Analysis Symposium, Portland, Maine, 3–6 October 2005.
28. Foody, G.M. Status of Land Cover Classification Accuracy Assessment. *Remote Sens. Environ.* **2002**, *80*, 185–201. [CrossRef]
29. Pontius, R.G., Jr.; Millones, M. Death to Kappa: Birth of Quantity Disagreement and Allocation Disagreement for Accuracy Assessment. *Int. J. Remote Sens.* **2011**, *32*, 4407–4429. [CrossRef]
30. Olofsson, P.; Foody, G.M.; Herold, M.; Stehman, S.V.; Woodcock, C.E.; Wulder, M.A. Good Practices for Estimating Area and Assessing Accuracy of Land Change. *Remote Sens. Environ.* **2014**, *148*, 42–57. [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.