Landscape Architecture

ARTICLE

Reliability-Aware Urban Green-Space Detectability from Sentinel-2 and OpenStreetMap Evidence in the Guangdong–Hong Kong–Macao Greater Bay Area

Kongjian Yu[1,*]

[1] College of Architecture and Landscape Architecture, Peking University, Beijing 100871, China

[*] Correspondence: kjyu@urban.pku.edu.cn

Abstract

The problem of urban green-space mapping in a highly concentrated metropolitan region involves consideration of the reliability of the decision itself rather than just the separation of vegetation from non-vegetation evidence. The Greater Bay Area of Guangdong-Hong Kong-Macao features high-rise neighborhoods, complex impervious surfaces, coastal towns, peri-urban vegetation, and an imbalanced amount of VGI data, indicating that the mere aggregation of accuracy measures will be inadequate for urban planning purposes. In this paper, it was examined whether reliability-weighted conformal graph calibration applied to the Sentinel-2 urban green-space evidence in combination with OpenStreetMap would be capable of discriminating among reliable mapped vegetation, city-level omission, and limitations imposed by structural connectivity. The reliability-weighted conformal graph calibration analysis used such inputs as the OpenStreetMap semantics for vegetation evidence, weight-sensitivity reliability scores, classifier confusion matrix numbers, city-level precision and recall scores, UGS area estimates, landscape metrics, and multi-scale partition analysis. The weighted reliability resulted in 0.8372 of mean accuracy and 0.8267 of mean F1 score, whereas unweighted setting decreased accuracy to 0.818 and F1 to 0.812. The best planning-oriented results of errors were observed for W-SVM with 355 false positives, 0.104 false positive rate, 0.896 specificity, and 0.785 utility. The detectability varied from 97.60 percent in Shenzhen to 67.42 percent in Kaiping, thus clearly showing a distinction between high-confidence cities and fragmentation-sensitive cities with considerable pressure on omission. Finally, the 10 m UGS estimation of 139,427.06 ha appeared to lie between ALCC, ESA WorldCover, and CLCD estimations, indicating substantial disagreements among urban green-space products. The MSPA analysis demonstrated that core and edge classes prevail in the foreground of vegetation evidence, whereas bridge, branch, and loop connectors combined account only for 3.11%.

Keywords: urban green space; Sentinel-2; OpenStreetMap; reliability weighting; conformal calibration; morphological spatial pattern analysis; Greater Bay Area

Received: 10 January 2025; Revised: 11 May 2025; Accepted: 21 September 2025; Published: 30 March 2026.

doi: 10.70517/la20261161-04

1. Introduction

Urban green space is both a type of land cover and an ecological resource. Within urban ecosystems, vegetation plays a role in regulating temperature, collecting rainwater, filtering air pollution, providing habitat pockets, and serving recreational and mental-health functions. All of these benefits depend on green-space distribution, which is characterized by parks, tree canopy, grassland, riparian vegetation, urban forestry, semi-natural hillslopes, and small green patches. Prior work has shown that the benefits of urban vegetation depend not just on the total coverage, but also on the spatial structure and accessibility of green assets [3, 13, 15, 30]. Furthermore, public-health studies and adaptation planning have emphasized the importance of urban natural environments in reducing thermal stress and improving environmental well-being [4, 31].

As a consequence of urbanization, satellite imagery is an indispensable resource for measuring and assessing urban green space. Sentinel-2 is a valuable tool for this application thanks to its ability to observe urban areas with 10–20 m resolution combined with spectral characteristics that measure vegetation vigour and moisture content in addition to built-up surface variability [10]. Indices such as the normalized difference vegetation index (NDVI), water content-based algorithms, and soil adjustment techniques continue to play an important role in identifying photosynthetic activity in urban settings, while spectral texture helps characterize local spatial structure [12, 20, 21, 23, 29, 34]. Even with all these tools, a 10 m optical product may fail to correctly identify green space in the compact districts where street trees, lawn, roof gardening, shaded parks, and small courtyards constitute only a portion of a single pixel.

Classification of urban land cover is similarly challenged by the mismatch between ecological objects and pixels. An urban park is an object that consists of internal paths, green grass, water, and canopy; a Sentinel-2 pixel is a spectral measurement that contains a subset of these components. The object-oriented nature of remote sensing has highlighted that spatial context is useful in mitigating some of the difficulties caused by spectral mixing; conversely, spatial pattern approaches have shown the importance of urban green-space shape and edge configuration in understanding land changes [2, 19]. In the Greater Bay Area, however, vertical land use exacerbates this problem. Tall buildings and viaducts cast shadows; roads and water boundaries create sharp edges; and urban green space is often linear and fragmented. Such a setting requires that reliability assessment be an intrinsic part of the analysis.

Volunteered geographic information (VGI) adds additional semantic value to urban classification efforts. OpenStreetMap (OSM) captures parks, forests, grassland, orchard, allotment garden, recreation grounds, scrub, vineyard, and many other land-use or natural features. These polygons can provide valuable labels for training supervised learning, which allows classifying green spaces within urban settings. The limitations of such data sources are well-known [14, 16, 18], as well: OSM completeness varies among cities, polygons contain heterogeneous internal surfaces, tags can differ in meaning among contributors, and geometries might be less accurate near edges and boundaries. For instance, an OSM polygon tagged as "forest" does not mean that all pixels within it will be correctly classified by the model. Rather, each polygon is a hypothesis that is used to filter candidate labels based on plausibility.

The implication for model design is direct: candidate labels need to be assessed for their spatial and spectral reliability. An internal forest polygon with dense canopy cover will not receive the same weight as the edge of a recreation ground with pavement inside it. Similarly, the urban setting with incomplete OSM contribution will not receive the same classifier weight as a city with many contributions and relatively homogeneous land cover. Reliability weighting is thus an instrument of data governance. It explicitly assigns the confidence level to each candidate label, which transforms the training evidence quality into an analytical element.

Thematic map accuracy assessment literature is rich in recommendations against assessing thematic maps by overall accuracy alone. Various types of accuracy, such as class-specific accuracies, precision, recall, error allocation, and area-based estimation, provide different perspectives on map quality [8, 11, 24, 26, 28]. The issue of accuracy assessment is particularly important in the case of urban green space since the consequences of overestimating or underestimating vegetation are different. False positive pixels indicate more abundant vegetation than it really exists, while false negatives may remove small parks and roadside vegetation from the map, thus giving a misleading impression of ecological provisioning. A credible map is therefore expected not only to perform adequately, but also to disclose where it is vulnerable.

The Greater Bay Area is a challenging setting for the problem at hand. Its urban system includes high-density central districts, manufacturing areas, estuarine edges, hill forests, peri-urban villages, and administration units with variable mapping completeness. At the same time, the region possesses significant amounts of ecological land that is subject to intensive development. This combination means that an identical classification algorithm may yield different results for various administrative units, which raises the issue of optimal classification threshold setting and introduces administrative diversity into the task. The mapping task is therefore multi-faceted, involving spectral and administrative aspects of the problem.

The central research question of this paper is whether it is possible to calibrate Sentinel-2 and OSM urban green-space evidence using reliability weighting to assess map reliability, class omission pressure, and ecological connectivity in the Greater Bay Area setting. The study builds upon the existing reliability-weighted conformal graph methodology, linking sample reliability, classifier performance, city detection, land cover area disagreement, and ecological connectivity in a coherent way.

Three specific questions will be answered. First, the study will clarify how OSM-derived weak labels should be understood as reliable information rather than uniformly correct training data. Second, the numerical stability, classifier utility, class omission pressure, and precision-recall imbalance will be compared between classifiers using values aggregated at the city level. Third, the mapping reliability will be linked to the landscape configuration in terms of area disagreement, patch structure, edge exposure, and lack of connectors. Overall, the study will show if reliability-based mapping is an appropriate approach for assessing urban green space.

2. Materials and Analytical Procedure

2.1. Metropolitan setting and mapping difficulty

The study focuses on the Guangdong–Hong Kong–Macao Greater Bay Area in southern China. It comprises Hong Kong and Macao along with Guangzhou, Shenzhen, Zhuhai, Foshan, Huizhou, Dongguan, Zhongshan, Jiangmen, Taishan, Kaiping, Enping, Heshan, Zhaoqing, and Sihui. The study region has an area of 5,600,000 ha and lies within the Pearl River Delta and its surroundings. It has a warm and humid subtropical climate and possesses rich vegetation, while being home to some of the densest urbanized areas in East Asia. Vegetation ranges from large hill forests to urban parks, road medians, riverbanks, orchards, recreation grounds, and small green spaces.

The study-region map shown in Figure 1 demonstrates the setting and presents four examples of imagery with problematic mapping conditions. Tall buildings cast shadows and reduce vegetation detection rates. Roadside vegetation may be too thin to be detected at 10 m resolution. Built-up and green spaces mixed within a single pixel create unreliable spectral signals. The boundary between urban vegetation and agricultural areas may be ambiguous. Such conditions call for a study that assesses classifier detectability and reliability.

This visual information, provided in Figure 1, clarifies why it is necessary for the calibration process to be sensitive to the local label reliability and administrative differences which influence the completeness of the OSM data.

(a) Greater Bay Area

(b) Local mapping conditions

Figure 1. Study region and mapping conditions.

The following section describes the data sets and methods used in the study.

2.2. Quantitative evidence and analytical variables

The quantitative evidence comprises the following list: OSM vegetation inventory, sensitivity value weighted by reliability, the accuracy result at the city level, classifier comparison counts, area of harmonized land cover, landscape pattern indices, and MSPA composition by Yuan et al. [36]. These values provide quantitative evidence for the evaluation of the reliability-weighted calibration without compromising the evidence base from Sentinel-2 and OSM data. The analytical variables are shown in Table 1.

Table 1. Evidence groups used in the analysis.

Evidence group	Analytical role
OSM vegetation inventory	Defines candidate green-space labels and clarifies which semantic classes enter the weak-label grammar.
Reliability-weight sensitivity	Tests whether weighted labels produce stable classification performance under alternative confidence values.
City-level performance	Supports detectability scoring, omission-pressure estimation, and precision–recall imbalance analysis by administrative unit.
Classifier comparison	Provides confusion-table counts and accuracy values for planning-sensitive utility analysis.
Land-cover area comparison	Quantifies agreement and disagreement among ALCC, ESA WorldCover, CLCD, and the 10 m UGS product after harmonization.
Landscape and MSPA values	Describes patch fragmentation, core dominance, edge exposure, and structural connector scarcity.

The evidence groups defined in Table 1 delimit the scope of the analysis, in which each quantitative statement is grounded, and interpretation follows Greater Bay Area logic. This approach ensures that the accuracy of maps, the consistency of the classified area, and the ecological connectivity of the product will not be considered separately; instead, multiple measures will be used to evaluate the same green-space product.

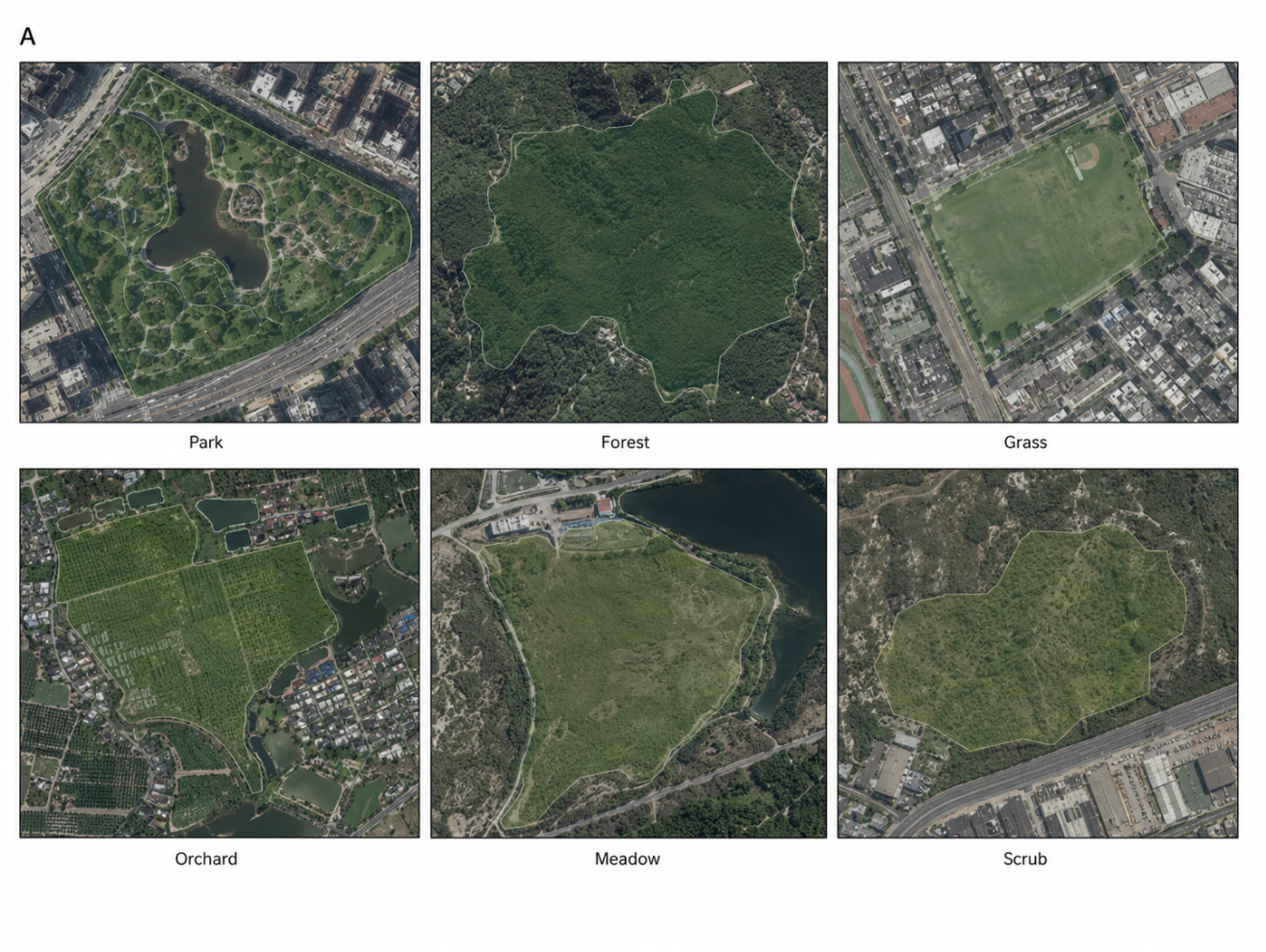
Data sources should be understood as a series of interrelated layers of evidence. Classification values reflect the quality of decisions made at the pixel level. The city values inform on the stability of the classifier across administrative units. The land cover comparison provides an indication of uncertainty at the area level. Landscape and MSPA values represent the transition from classification results to ecology through a transformation of the binary map. Such an approach helps to avoid the limitation of many urban green-space analyses, namely, the presentation of high classification value without information about spatial continuity, administrative balance, or product stability.

The OSM label grammar is defined in Table 2. The categories are considered as approximations, rather than an absolute truth. Parks can have paved paths and water; recreational grounds can combine grass with sports surfaces; orchards or vineyards are vegetation areas, but not urban green spaces. These differences in meaning are precisely what calls for a reliability-weighting procedure.

As one can see in the example images in Figure 2, OSM polygons that have green interiors with homogeneous vegetation indicators tend to be more reliable compared to small borders, highly heterogeneous surfaces, or shaded urban areas. Dependability surface incorporates this principle into the very definition of candidate labels, dividing dependable interiors from unreliable borders and heterogeneous urban classes.

The distribution of classes in the database according to Table 2 explains the increased complexity of the database in comparison with a binary vegetation mask. The class "candidate UGS" is quite general in terms of meaning, and its

inner structure varies from city to city. Spectrally and geometrically different objects (forest, grass, scrub, parks) cannot be equally reliable.



(a) OSM semantic examples

(b) Reliability surface

Figure 2. Weak-label reliability.

Table 2. OSM categories for candidate green-space labels.

Key	Value	Vegetation type	Interpretation
Landuse	Allotments	Crop	Small cultivated garden areas with mixed vegetation.
Landuse	Farm	Crop	Agricultural land with planted areas, trees, and farm structures.
Landuse	Forest	Forest	Woodland or forest areas with continuous canopy.
Landuse	Grass	Grass	Grass-dominated surfaces, including lawns and open grass.
Natural	Heath	Shrub	Low shrub or heath vegetation.
Landuse	Meadow	Grass	Meadow or grazing vegetation with seasonal variation.
Leisure	Nature reserve	Mixed	Protected vegetation with mixed natural cover.
Landuse	Orchard	Forest	Tree-crop vegetation with canopy structure.
Leisure	Park	Mixed	Urban or peri-urban park space with internal mixed surfaces.
Landuse	Recreation ground	Grass	Open recreation areas, commonly grass-dominated but not always pure vegetation.
Natural	Scrub	Shrub	Scrub or woody shrub vegetation.
Landuse	Vineyard	Forest	Cultivated vine vegetation treated as vegetation-related cover.

2.3. Reliability-weighted conformal graph calibration

The method used for this purpose is Reliability-Weighted Conformal Graph Calibration. The sample set for calibration includes candidate samples with an index of $i = 1, \ldots, n$, in which the feature vector is denoted by x_i, class indicator $y_i \in \{0, 1\}$, and label reliability by $r_i \in 0, 1$. Feature vector may include visible, near-infrared, red-edge, and shortwave infrared bands, vegetation and water indexes, as well as texture indicators. It is necessary because urban green surfaces are different in terms of canopy height, phenology, shading, moisture, and proximity to buildings. Reliability is represented as a continuous score:

$$r_i = \text{clip}\left(\beta_0 \;\; \beta_s S_i \;\; \beta_g G_i \;\; \beta_c C_i - \beta_b B_i - \beta_m M_i, 0, 1\right), \tag{1}$$

where S_i is spectral plausibility, G_i is geometric stability, C_i is city-level completeness, B_i is boundary risk, and M_i is a mixed-shadow penalty. Spectral plausibility protects against paved or water-dominated pixels inside green polygons. Geometric stability favors polygon interiors over edges. City completeness prevents under-mapped administrative units from being treated as equally reliable. Boundary and mixed-shadow terms reduce the confidence of pixels most likely to cause false positives or omissions.

The classification loss is weighted by reliability:

$$\mathcal{L} = \sum_{i=1}^{n} r_i \ell y_i, p_i \;\; \lambda_\Omega \sum_{m=1}^{M} \Omega f_m, \tag{2}$$

where $p_i = \sum_{m=1}^{M} \omega_m f_m x_i$ is the ensemble probability, $\ell y_i, p_i$ is binary cross-entropy, and Ωf_m is learner-specific regularization. The classifier comparison includes ANN, RF, XGBoost, and W-SVM; these model families are established in remote sensing and machine-learning literature [1, 6, 7, 9]. Weighted support vector classification is especially relevant because margin-based separation can be adjusted by sample reliability, and probability calibration can translate model scores into more interpretable decision confidence [25].

A city-specific conformal gate controls uncertain predictions:

$$\hat{z}_i = \begin{cases} 1, & p_i \geq \tau_c, \\ 0, & p_i \leq 1 - \tau_c, \\ u, & 1 - \tau_c < p_i < \tau_c, \end{cases} \tag{3}$$

where τ_c is the city-level threshold and u denotes uncertainty. Conformal prediction is useful here because it uses calibration behavior to regulate the confidence of predictive sets [33]. In planning terms, the uncertain state is valuable because it avoids presenting ambiguous mixed pixels as definite green-space supply.

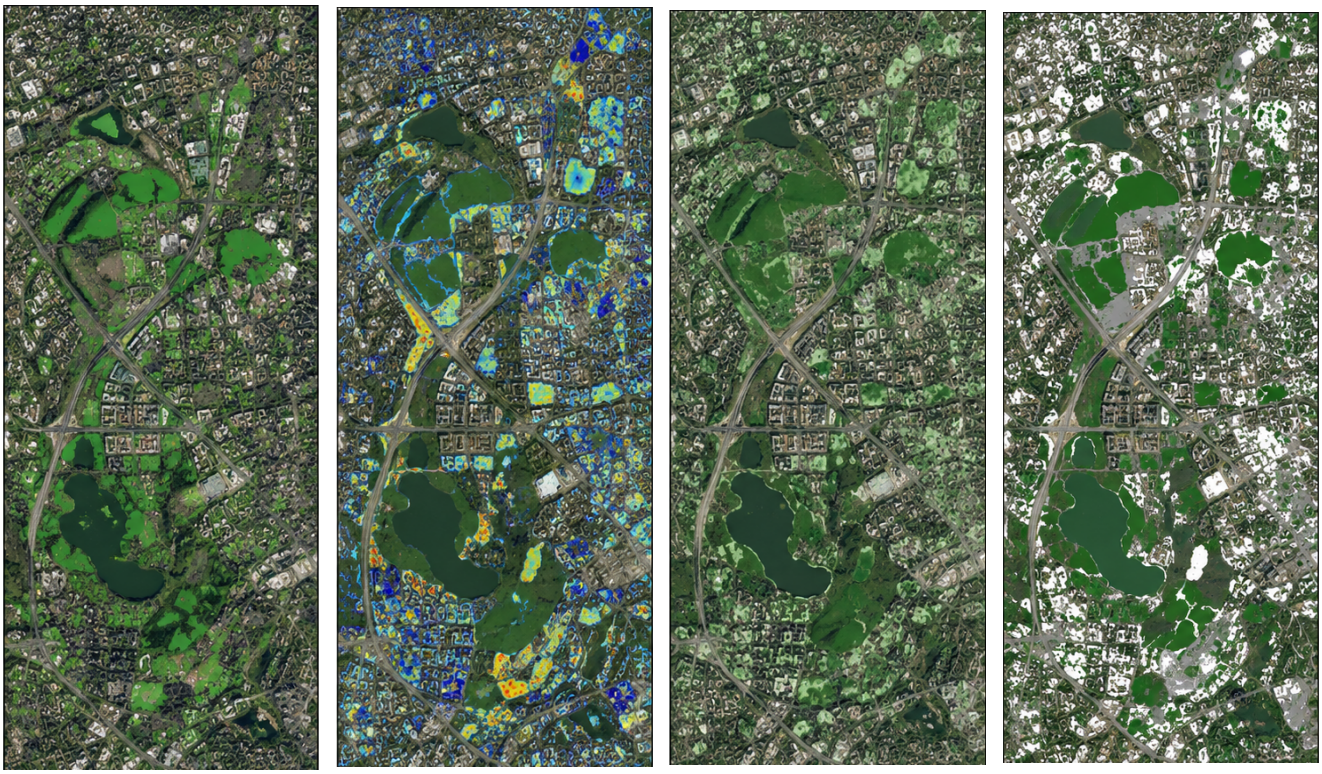
Spatial refinement uses a graph energy:

$$Ez = {}_i \phi_i z_i \;\; \lambda_G \, {}_{i,j\in\mathcal{E}} w_{ij} \mathbb{I} z_i \neq z_j, \tag{4}$$

with

$$w_{ij} = \exp\left(-\frac{\|v_i - v_j\|^2}{\sigma_v^2} - \frac{d_{ij}^2}{\sigma_d^2} \right). \tag{5}$$

Here, $\phi_i z_i$ is derived from calibrated probability, $\mathcal{E}$ is the adjacency set, v_i and v_j are spectral–textural descriptors, and d_{ij} is spatial distance. Graph-based energy minimization is a standard way to combine local evidence and spatial consistency [5]. The graph term is interpreted conservatively. It suppresses isolated salt-and-pepper decisions only when neighboring pixels are spectrally compatible, so narrow but coherent vegetation is less likely to be removed.



(a) Candidate labels **(b)** Reliability **(c)** Probability **(d)** Refined decision

Figure 3. Pixel decision states.

The four panels in Figure 3 translate the equations into a local decision sequence. Candidate labels define where green-space evidence begins; reliability distinguishes strong and weak label support; calibrated probability expresses

model confidence; graph refinement removes isolated noise while keeping coherent green patches. The sequence emphasizes progressive evidence control at pixel scale.

The practical value of this sequence is that it separates three decisions that are often collapsed into one. A pixel can belong to a candidate polygon but still be unreliable. It can have a moderate vegetation probability but still fall inside the conformal uncertainty band. It can be locally isolated but still be retained if neighboring spectral evidence supports a narrow green feature. This separation is important for dense urban vegetation because many planning-relevant features are small, linear, or partially shaded.

2.4. Diagnostic indices

The analysis uses city-level and landscape-level measures to interpret the classification output. City detectability is defined as

$$D_c = 1000.30A_c \;\; 0.25P_c \;\; 0.25R_c \;\; 0.20F_c, \tag{6}$$

where A_c, P_c, R_c, and F_c are accuracy, precision, recall, and F1 score. Omission pressure and precision–recall imbalance are

$$O_c = 1 - R_c, \tag{7}$$

$$B_c = |P_c - R_c|. \tag{8}$$

These equations turn standard accuracy values into planning information. High detectability indicates that a city's mapped green-space decisions are balanced across multiple performance indices. High omission pressure indicates that vegetation may be missed. High imbalance indicates that the model behaves asymmetrically and may need city-specific threshold control.

Area consistency is evaluated using

$$S_k = \frac{U_k}{U_k \;\; N_k}, \tag{9}$$

$$\Delta_k = 100\frac{U_k - U_{10m}}{U_{10m}}, \tag{10}$$

where U_k and N_k are UGS and non-UGS areas for product k, and U_{10m} is the 10 m UGS estimate. The equations allow the 10 m product to be compared with ALCC, ESA WorldCover, and CLCD while recognizing that global and national land-cover products differ in class definitions, spatial resolution, and temporal compositing [35, 37].

Landscape diagnosis uses patch and morphological spatial pattern concepts. Spatial pattern indices have long been used to interpret the form of land-cover patches and urban growth [2, 19, 22]. MSPA interprets binary foreground patterns as core, edge, bridge, branch, loop, islet, and perforation components [27, 32]. Connector scarcity and edge exposure are represented by

$$C_{mspa} = \frac{\text{Bridge} \;\; \text{Branch} \;\; \text{Loop}}{\text{Core}}, \tag{11}$$

$$E_{mspa} = \frac{\text{Edge}}{\text{Core}}. \tag{12}$$

These indicators connect map classification to ecological interpretation. A product can achieve reasonable pixel accuracy while still showing a spatial system dominated by exposed edges and weak connectors.

3. Results

3.1. Reliability weighting stability

Reliability weighting produced stable performance across the tested confidence settings. Across the nine weighted schemes, mean accuracy was 0.8372, mean precision was 0.8851, mean recall was 0.7754, and mean F1 score was 0.8267. The standard deviations were small, with 0.0008 for accuracy and 0.0016 for F1 score. The unweighted setting reduced accuracy to 0.818 and F1 score to 0.812. This result shows that reliability information is not cosmetic; it gives heterogeneous OSM labels differentiated influence while remaining insensitive to minor changes in the high-, medium-, and low-confidence values.

The evidence in Table 3 supports two interpretations. First, the weighted configurations are tightly clustered, which suggests that the procedure depends on the presence of reliability differentiation rather than on a fragile numerical tuning. Second, the unweighted case performs worse even though its recall is close to some weighted schemes. That pattern means the loss of weighting mainly weakens precision and balance, not simply the ability to find green pixels. For urban planning, that distinction matters because overstating green-space area can be as problematic as missing small fragments.

Table 3. Reliability-weight sensitivity.

Scheme	W_{high}	W_{medium}	W_{low}	Accuracy	Precision	Recall	F1
S01	1.00	0.60	0.20	0.836	0.885	0.772	0.825
S02	0.90	0.60	0.30	0.838	0.884	0.779	0.828
S03	1.00	0.70	0.40	0.838	0.882	0.779	0.828
S04	1.00	0.50	0.10	0.837	0.891	0.769	0.825
S05	1.00	0.60	0.30	0.836	0.885	0.773	0.825
S06	0.95	0.55	0.15	0.837	0.888	0.772	0.826
S07	0.85	0.55	0.25	0.838	0.884	0.778	0.828
S08	0.90	0.50	0.10	0.837	0.889	0.771	0.826
S09	1.00	0.80	0.40	0.838	0.878	0.786	0.829
S10	1.00	1.00	1.00	0.818	0.858	0.778	0.812

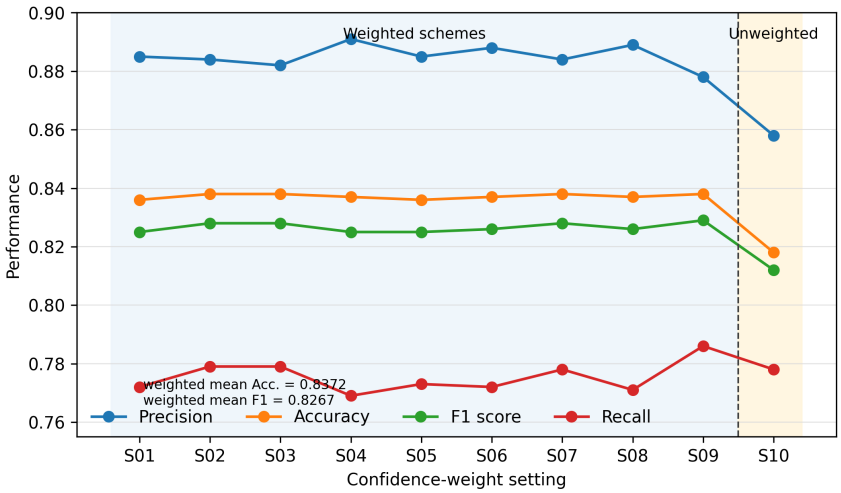


Figure 4. Weighted-scheme stability.

The plot in Figure 4 allows one to understand the results presented in the table better. Both accuracy and F1 score are nearly constant when changing from S01 to S09, but S10 diverges from the weighted group. It is obvious from the table that precision is greater than recall for all weight sets except S10. This behavior will be explained below in terms of how city-level omission bias emerges and becomes important.

This result reinforces the explanation above because the choice of any specific triple of weights does not play any role here. Otherwise, the method required more fine tuning and would become less transferable from one city to another. In other words, we have no strong dependency on the way that we use to select our training samples. That is good news for urban planners who need a simple and robust approach to map UGS.

3.2. Classifier error control

When comparing the performance of classifiers, it is wrong to pay attention only to recall. Recall is the best among all compared methods in RF, which gave 551 false positives. The lowest number of false positives was in W-SVM, whose specificity was also the highest one. W-SVM got the best utility score at 0.785 when false positives were considered two times as harmful as false negatives. At the same time, RF, XGBoost, and ANN provided values of 0.733, 0.727, and 0.715, respectively.

The values in Table 4 clarify the error economy of the applied method. RF retrieved more green-space pixels, but its higher false-positive count would inflate green-space supply in a planning map. W-SVM did not maximize recall, yet it offered the most credible balance because it strongly reduced false positive assignment while maintaining acceptable recall. The result justifies the conformal gate in the applied procedure: uncertain probabilities should be regulated before they are converted into final green-space labels.

Table 4. Classifier utility comparison.

Classifier	Acc.	Prec.	Rec.	F1	TN	FP	FN	TP	FPR	FNR	Spec.	Utility
ANN	0.797	0.822	0.761	0.791	2819	563	816	2605	0.166	0.239	0.834	0.715
RF	0.814	0.835	0.797	0.815	2751	551	712	2789	0.167	0.203	0.833	0.733
XGBoost	0.804	0.837	0.770	0.802	2777	525	805	2696	0.159	0.230	0.841	0.727
W-SVM	0.837	0.882	0.779	0.827	3047	355	753	2648	0.104	0.221	0.896	0.785

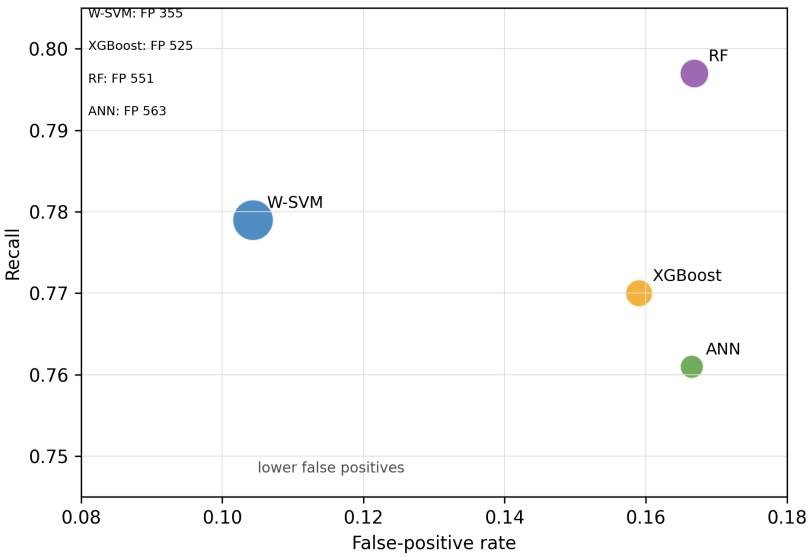


Figure 5. Classifier error-cost space.

Figure 5 illustrates how W-SVM is the model nearest to the preferable planning region due to the combination of both low FPR value and high utility. ANN and RF represent different kinds of weaknesses: ANN displays low recall as well as a relatively high number of FP, and RF enhances recall at the expense of a loss in FP management. XGBoost demonstrates an intermediary behavior. This implies that we should interpret Figure 5 from the planning standpoint, not as a model-ranking task focused only on accuracy.

A false-positive focus does not mean that the issue of omissions can be overlooked since there is another risk for a green-space product to be considered useful. False positives can exaggerate green-space resources and obscure any lack thereof in densely populated areas. False negatives can make it necessary to eliminate small but real green-space assets. Here the conformal gate and city-level diagnostics play their roles: the former makes sure that too confident assignment does not occur, and the latter shows which cities suffer from excessive conservative detection.

3.3. Urban green-space detectability by city

There are significant variations in administrative heterogeneity detected using city-level diagnostics. The highest detectability score of 97.60 was achieved by Shenzhen, with 0.976 accuracy, 0.980 precision, 0.972 recall, and 0.976 F1 score. Foshan, Enping, and Zhaoqing cities made up the stable group of high-confidence scores. Kaiping received the lowest detectability score of 67.42, with 0.467 recall and omission pressure of 0.533. Fragment-sensitive cities include Heshan, Zhongshan, and Guangzhou. The mean detectability score equals 82.30 with a standard deviation of 6.21. Mean precision reaches 0.885, but recall is 0.759.

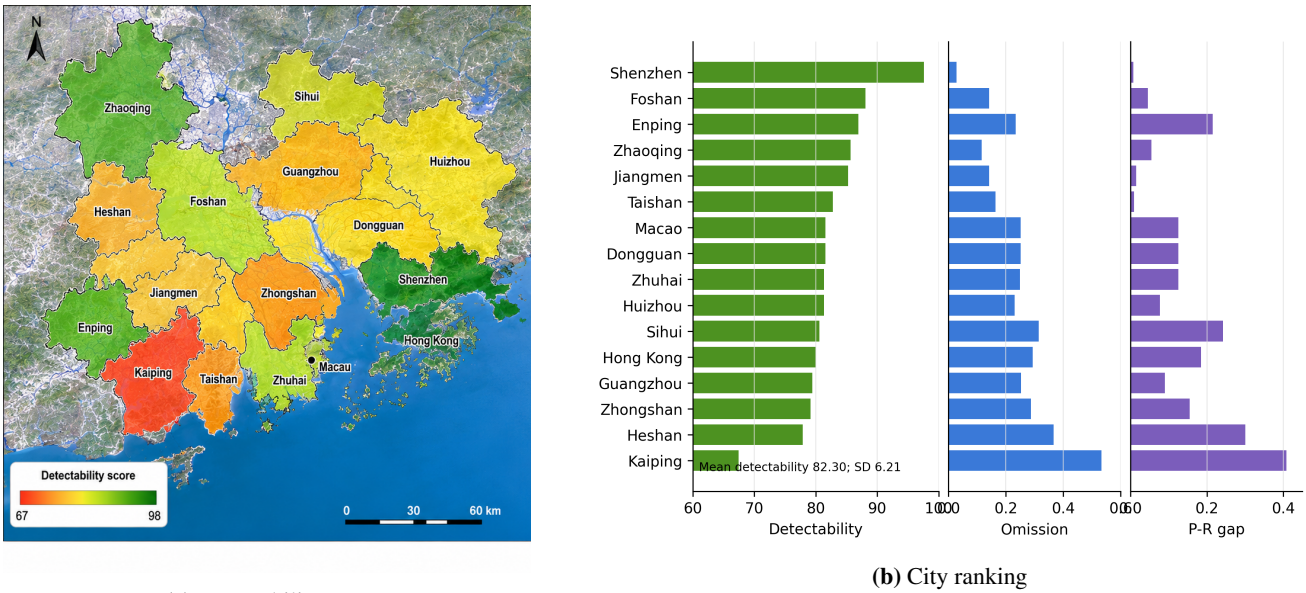
Table 5. City detectability diagnostics.

City	Acc.	Prec.	Rec.	F1	D_c	O_c	Diagnostic class
Shenzhen	0.976	0.980	0.972	0.976	97.600	0.028	Stable high-confidence
Foshan	0.879	0.905	0.859	0.881	88.090	0.141	Stable high-confidence
Enping	0.868	0.981	0.766	0.861	86.935	0.234	Stable high-confidence
Zhaoqing	0.858	0.829	0.884	0.855	85.665	0.116	Stable high-confidence
Jiangmen	0.857	0.843	0.858	0.851	85.255	0.142	Moderate heterogeneous
Taishan	0.819	0.827	0.836	0.832	82.785	0.164	Moderate heterogeneous
Macao	0.832	0.873	0.748	0.805	81.585	0.252	Moderate heterogeneous
Dongguan	0.832	0.873	0.748	0.805	81.585	0.252	Moderate heterogeneous
Zhuhai	0.817	0.876	0.751	0.809	81.365	0.249	Moderate heterogeneous
Huizhou	0.829	0.846	0.769	0.805	81.345	0.231	Moderate heterogeneous
Sihui	0.815	0.928	0.686	0.789	80.580	0.314	Moderate heterogeneous
Hong Kong	0.810	0.892	0.707	0.789	80.055	0.293	Moderate heterogeneous
Guangzhou	0.802	0.837	0.747	0.789	79.440	0.253	Fragment-sensitive
Zhongshan	0.799	0.868	0.713	0.783	79.155	0.287	Fragment-sensitive
Heshan	0.788	0.933	0.633	0.754	77.870	0.367	Fragment-sensitive
Kaiping	0.723	0.875	0.467	0.609	67.420	0.533	Fragment-sensitive

The city results in Table 5 show that reliability is not spatially uniform across the Greater Bay Area. High precision and low recall in several cities indicate that the product is often conservative: it avoids many false positives but can miss actual green space. Kaiping and Heshan are the clearest examples, because their precision remains relatively high while recall is substantially lower. For these cities, the problem is not random error but systematic omission of

difficult green-space forms.

In the city atlas of Figure 6, there is a spatial reading of the city table, where the stable high-confidence cities stand apart from the fragment-sensitive ones, and it is apparent in the ranking panel that omission pressure increases with decreasing detectability. Such results imply city-wise calibration instead of one regional threshold. At the same time, the output suggests why a map can be trusted on an overall basis but will still require additional verification in particular administrative divisions.



(a) Detectability map

(b) City ranking

Figure 6. City-level reliability pattern.

When looking at both precision and recall, one may find that although recall is relatively low for some cities, the precision of several weaker cities implies that the majority of mapped green pixels can be considered correct. However, recall's values show that the product under study is probably incomplete since smaller, mixed or mislabelled green areas are possibly unrepresented in the dataset. Therefore, in terms of practical use, planning must incorporate a mapped inventory with the omission pressure alert.

3.4. Product agreement in mapped green-space area

From a perspective of harmonized area comparison, the 10 m UGS dataset falls within the estimates made by other products. According to ALCC's data, there are 97,668.27 ha of urban green spaces, which constitute a 30.0% underestimation compared to the 10 m UGS data. By contrast, according to ESA WorldCover, the area of UGS amounts to 125,108.40 ha, 10.3% less than in the 10 m product. The estimate produced by CLCD equals 183,620.34 ha, 31.7% higher than the 10 m UGS data. The 10 m UGS product contains 139,427.06 ha of green spaces, implying a share of 0.212.

Table 6. Harmonized UGS area comparison.

Dataset	UGS area (ha)	Non-UGS area (ha)	UGS share	Deviation
ALCC 30 m	97,668.27	556,316.46	0.149	-30.0%
ESA WorldCover 10 m	125,108.40	533,790.82	0.190	-10.3%
CLCD 30 m	183,620.34	471,589.65	0.280	31.7%
10 m UGS product	139,427.06	519,503.01	0.212	0.0%

The area values in Table 6 show that product choice can change regional UGS estimates by more than 30% in either direction. This magnitude is large enough to affect planning conclusions about green-space sufficiency. The 10 m UGS product does not sit at an extreme; it lies between CLCD and ALCC and closer to ESA WorldCover. That position supports its plausibility, while the spread among products still requires cautious interpretation.

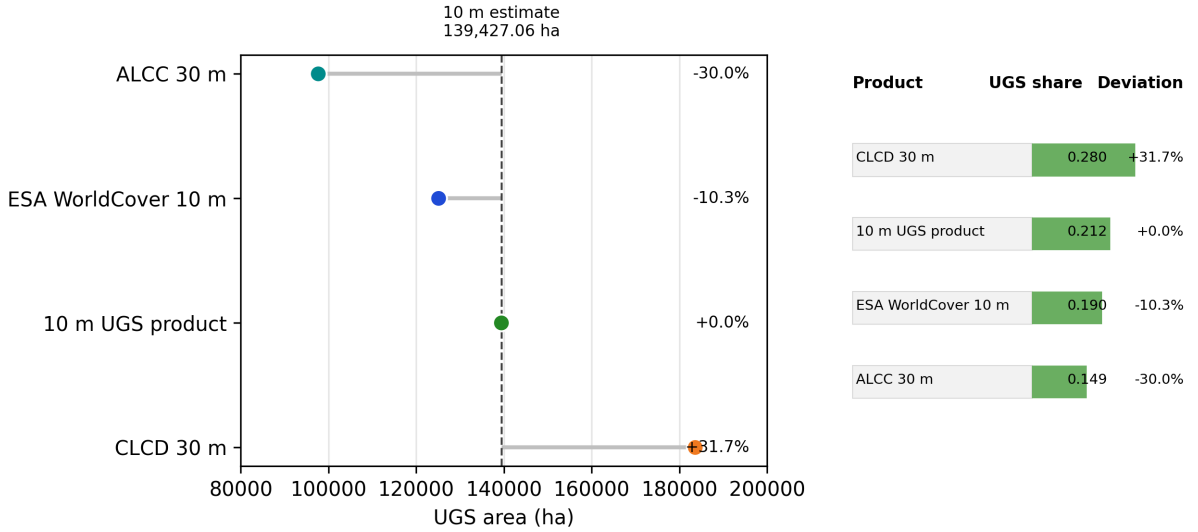


Figure 7. UGS area envelope.

However, this envelope can be used to make the discrepancy visible without the implication that one external product is the only right choice. Cropland, grassland, shrub, forest, wetland, and urban fringe vegetation can be classified quite differently using various land cover products. Densely-populated areas will be under-represented in dense districts due to coarse-resolution; large categories of vegetation classes in peri-urban areas might result in an overestimate of the green space. Therefore, the 10 m estimate must be interpreted within a product-related range of uncertainty.

The area envelope will also prevent any exaggeration of differences among cities and products. Policy statements that rely on area estimates from one city or one product may seem quite accurate, even though definition criteria for classifications are different. Thus, by placing the 10 m estimate among other lower and higher estimates, one can obtain a more solid ground for comparing results. Moreover, it also illustrates why reliability should be coupled with area: despite a credible estimate of the UGS area, further local verification is needed if reliability is poor or vegetation mapping exhibits structural fragmentation.

3.5. Patch configuration and MSPA structure

Fragmentation parameters indicate fragmented landscape structure. At the GBA scale, the numbers of patches were 88,635, largest patch index equaled 0.07, landscape shape index – 366.32, and the area-weighted mean patch fractal dimension was 1.14. Large cities possessed an even greater amount of patches, i.e., 104,446 and had largest patch index equal to 0.06, thus illustrating a very fine-grained structure of the landscape. Small cities showed less continuous geometry of vegetation cover by demonstrating 4,312 patches, largest patch index of 0.23, and landscape shape index of 77.85. However, parks had the highest patch index, which equaled 0.41, along with cohesion of 93.20, but the agglomeration index of 87.08 did not exceed 90.21.

The pattern of the landscape configuration presented in Table 7 illustrates that green-space amount and green-space pattern should be analyzed together. Big cities have lots of green-space patches but low domination by any big

patch in them; this corresponds to high fragmentation. Small cities have fewer green-space patches and high largest patch index; therefore, green surfaces are not so fragmented and are rather connected. Park systems are internally coherent structures, but relatively low agglomeration proves the lack of coherence between park systems and other elements of green space structure.

Table 7. Landscape configuration values.

System	NP	LPI	LSI	COHESION	AI	FRAC_AM
GBA overall	88,635	0.07	366.32	–	90.21	1.14
Large cities	104,446	0.06	375.56	–	89.10	1.14
Small cities	4,312	0.23	77.85	–	85.82	1.13
Urban parks	–	0.41	–	93.20	87.08	–

As for structural connectivity of the landscape pattern, it can be estimated using MSPA values. For example, core made up 56.38% of the green structure area, while edge occupied 39.19%. Altogether core and edge covered 95.57% of green space. Bridges, branches, and loops made up only 3.11% of the mapped structure. The ratio between edges and cores amounted to 0.695, and the ratio between core and connectors was 18.13. Thus, the regional green-space system has lots of cores and edges but lacks structural connectors.

The MSPA values in Table 8 define the ecological interpretation. The problem is not simply that the region lacks core vegetation. Instead, the core-dominated pattern is accompanied by a large edge fraction and very limited connector elements. Mapping reliability must therefore be interpreted with landscape structure: a map can correctly identify large green surfaces while still revealing weak ecological continuity.

Table 8. MSPA composition and indicators.

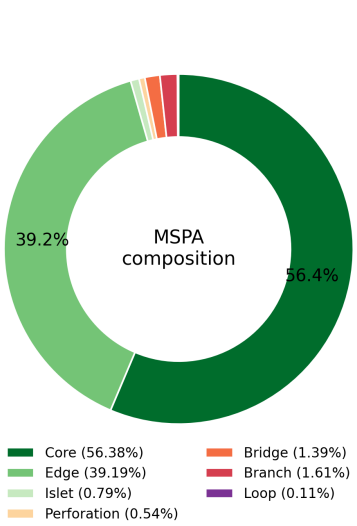
Component or indicator	Value	Interpretation
Core	56.38%	Large internal green blocks remain important.
Edge	39.19%	Boundary exposure is high.
Islet	0.79%	Small isolated patches have limited area share.
Perforation	0.54%	Internal openings are present but not dominant.
Bridge	1.39%	Direct connectors are scarce.
Branch	1.61%	Dead-end connectors are limited.
Loop	0.11%	Redundant circular connections are negligible.
Bridge + branch + loop	3.11%	Connector fraction is low.
Edge/core ratio	0.695	Edge exposure is substantial.
Core/connector ratio	18.13	Connectivity is the main structural constraint.

The morphological plate in Figure 8 emphasizes the disparity between abundant core structures and few connectors. The prevalence of core and edge suggests that the area maintains its vegetated biomass, yet the scarcity of connectors restricts connections and redundancy. This insight leads us away from an interpretation based merely on the increase in green space area and toward one based on corridors, edge protection, and local connector rehabilitation.

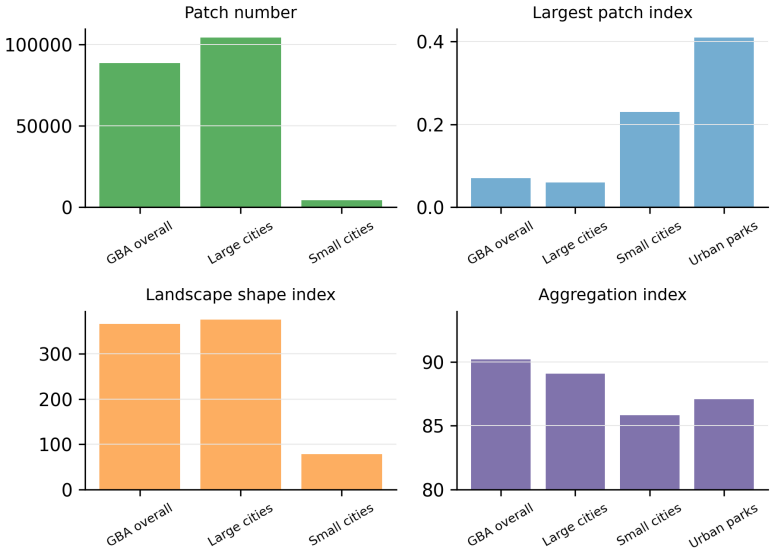
This point is crucial for ecological analysis. If the focus were simply on increasing the green space area, the area in question would receive protection, while the areas between would be degraded. A lack of connectors would create a weak system, as connectivity in nature depends on the links between larger patches. The MSPA approach thus provides the classification algorithm with another function—it shows not only where green biomass can be found, but also how it is configured.

The synthesis in Figure 9 translates the numerical results into recommendations for urban green-space planning.

High-confidence stable cities can afford routine monitoring and core protection, whereas fragment-sensitive cities require omission control, confidence thresholds reevaluation, and focused validation in dense or heterogeneous urban conditions. On the regional level, the connector fraction of 3.11%, the edge/core ratio of 0.695, and the core/connector ratio of 18.13 suggest that strengthening the core corridors and stabilizing the edges is necessary to complement classification reliability.



(a) MSPA composition



(b) Landscape indices

Figure 8. Green-space morphology.

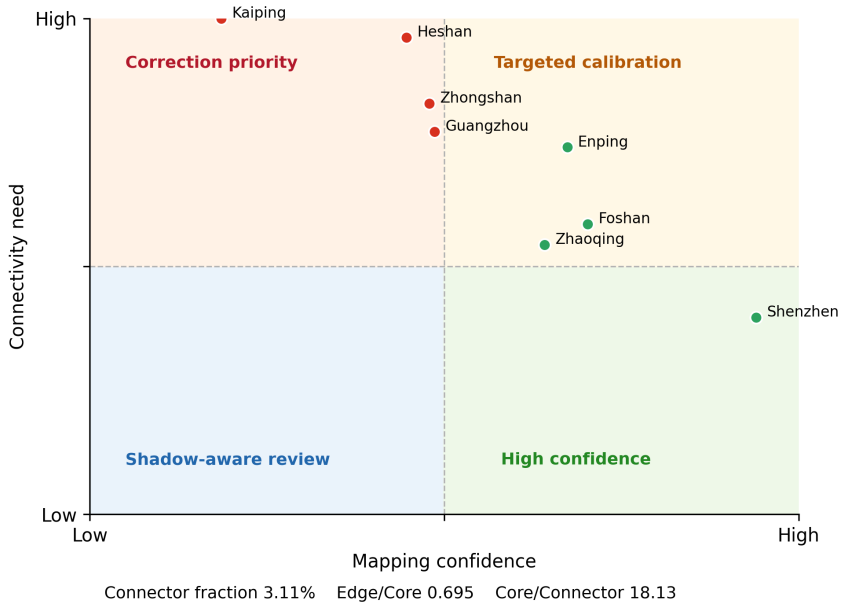


Figure 9. Planning translation.

The spatial reliability diagnostic ensures that a single solution will not work everywhere within the region. When the conditions are like those of Shenzhen, the monitoring and maintenance strategy applies because the map has high credibility. For conditions similar to Kaiping, validation and omission control are called for because the map has less complete coverage. Moreover, morphological characteristics add yet another condition: even where confidence is high, the rarity of connectors requires planning. In other words, the greatest usefulness of the map is diagnostic.

4. Discussion

4.1. Label reliability and map credibility

From the results, it appears that weak-label reliability is a crucial issue in the credibility of the greater Bay Area's urban green-space mapping using OSM and Sentinel-2. Although OSM polygons yield semantically meaningful evidence, their classification is not based on homogeneous reference data. Parks, recreational sites, orchards, scrub vegetation, grassland, and forests represent different combinations of internal structure, geometry, and spectral properties. Weighting schemes were found more effective than the non-weighted variant of classification, and they proved remarkably stable regardless of confidence levels. It can be concluded that the main contribution does not lie in numerical tuning of the weights, but in their recognition as indicators of label quality.

The reliability score captures both aspects of data quality. Spectral plausibility discounts the evidence for pixels whose label appears spectrally implausible. Geometrical stability distinguishes interior polygons from boundaries that are prone to variability. Completeness rules out the possibility of using administrative variations in OSM data as a means to increase classification accuracy. The mixed-shadow penalty addresses an important source of error in compact high-rise cities. All these features make up a comprehensive effort at improving the truthfulness of the training evidence.

In the context of joint analysis of remote-sensing imagery and VGI, this approach becomes even more significant. The latter tends to produce highly valuable labels thanks to the richness of semantics, but this very feature may introduce unwanted diversity. The application of reliability-aware design transforms the variability into an explicit form of evidence. Instead of presenting the map as perfectly observed reality, it is presented as an evidence product in which some labels, some pixels, and some cities are more reliable than others. This honesty pays off when applying the map for decision-making.

4.2. False-positive control for planning use

Classification utility proves that a UGS map designed for planning must control for false positives. A planner should avoid any map-based estimation of urban green-space supply that may inflate the actual supply in dense neighborhoods. While RF classifier yielded a higher recall score than W-SVM, the number of false positives was considerably greater in RF maps. W-SVM decreased false positives to 355, making its specificity of 0.896 higher than the non-calibrated counterpart. For these reasons, W-SVM seems superior for a green-space map used for planning purposes.

The conformal gating method immediately follows from these observations about the utility scores. As opposed to forced probabilistic labeling in non-conformal classifiers, the conformal gate leaves the uncertain states open to further examination. This step is particularly critical in cities characterized by large imbalance between precision and recall. In these cases, a binary choice between non-green and green pixels yields either inflated or underestimated supply of urban green spaces.

4.3. Administrative heterogeneity and city-specific calibration

According to the city diagnostics, the greater Bay Area should not be regarded as an administrative entity for mapping purposes. Shenzhen represents a city with high accuracy, high precision, recall, and F1 score, meaning that its class evidence and calibration are strong. The opposite is true about Kaiping, where the omission pressure

is very high and the detectability score is low. Heshan, for its part, displays a large discrepancy between precision and recall scores. All this suggests that there are cities with more cautious approaches to vegetation classification.

City-specific calibration is thus not only a practical convenience, but a necessity dictated by the administrative situation. Highly compact and fragment-sensitive cities require more attention devoted to uncertain evidence and validation efforts. Cities characterized by high stability can rely on routine monitoring, while moderate cases call for more thorough analysis of the reasons for omission. Is the problem of missing classes due to shadows, small size of vegetation patches, incomplete OSM data, or general uncertainty about urban land cover? This approach is more informative than regional averaging.

4.4. Area disagreement and planning uncertainty

The land cover comparison demonstrates that the harmonized UGS area strongly depends on the land cover product. ALCC was lower than the 10 m UGS estimate, CLCD was higher, and ESA WorldCover was closer to the latter but still below. This difference was expected since different land cover products use various classification criteria, spatial resolutions, composite images, and minimum mapping units. The UGS classes were particularly sensitive to these factors since the green space may be small, fragmented, or located near the boundary of urban fringe.

The area envelope should be understood as a range of uncertainty of the product. 139,427.06 ha is the number belonging to a set of alternative estimates of UGS area. This understanding is useful for environmental governance since, in deciding about green space sufficiency in a particular city, the policy maker must take this uncertainty into consideration. For example, even a city with relatively high area estimate and low omission pressure may require verification to establish the actual green space presence.

4.5. Connectivity constraints in core-dominated morphology

The landscape analysis and MSPA statistics demonstrate that the problem of urban green-space area is complemented by a connectivity problem in the greater Bay Area. The share of core in the foreground exceeds 50%, and the proportion of edge vegetation reaches almost 40%. Connectors are exceedingly scarce. This suggests that the ecological functions may be limited by poor connectivity rather than the lack of area of green cover. Large green spots will be visible on the map, but connecting corridors will be missing.

The edge/core ratio and the core/connector ratio imply two kinds of intervention priorities. High edge exposure means the necessity of buffering urbanization, controlling encroachment and managing park boundaries. High connector scarcity implies the necessity of riverbank restoration, greening of roadways and transport corridors, coastal belt preservation, and green small fragments linking cores. Compact districts will also include the establishment of pocket parks, courtyard gardens, school greens, institutional trees, and even roof and podium vegetation.

The same interpretation also contributes to our understanding of the interaction between mapping uncertainty and landscape planning. Fragment-sensitive cities must improve vegetation detection, but they may contain exactly the fragments needed for connectivity. If they are missed in the map, then the plans of corridors will overlook stepping stones already present in urban landscape. Conversely, mixed impervious surfaces incorrectly classified as vegetation will serve as false connections. Hence, reliability-aware mapping is a precondition of reliable green corridor designs.

4.6. Validation requirements

The findings of the research depend on the interpretation of numerical values rather than re-analysis of the pixel stack. Nevertheless, they provide detailed information regarding the performance of reliability-aware calibration of the classification algorithm. Full validation would require re-processing of Sentinel-2 image composites and OSM vector data, taking into account administrative divisions, calibration samples, and scripts for feature extraction, reliability scoring, conformal gating, and graph optimization.

There are several issues to be resolved regarding validation. First, the test sample must be spatially independent to prevent optimistic estimates caused by validation clusters. Second, verification in dense districts should use high-resolution remote-sensing imagery or field observation. Third, OSM completeness must be quantified separately for each city in order not to derive city-level reliability solely from classification. Fourth, the conformal gate should be tested across urban cores, transition peri-urban zones, and ecological mountains.

Operational use of the classification output will require presentation of uncertainty in an interpretable form. Planning agencies will not need every probability value but will want to know districts with omission problem, weakly developed corridors, and sensitivity of area estimate. Calibration may be used in such cases in exporting three parallel products – a green space class map, a reliability or uncertainty layer, and a connectivity diagnostic layer. This will provide classification evidence for park planning, ecological restoration, and monitoring of metropolitan environment.

Finally, there is no reason to expect that results for the greater Bay Area can be generalized for other regions. This area has humid subtropical vegetation, intensive vertical construction, and dense mixture of formal and informal green spaces. Other places may display quite a different relationship between precision, recall, and omission depending on phenology, building density, and vegetation inventories. Instead of trying to use the same thresholds in another metropolitan area, one should recognize the transferable element – the evaluation logic. Label reliability, uncertain prediction, and landscape structure should be considered together.

5. Conclusion

Reliability-weighted calibration of urban green-space evidence based on Sentinel-2 and OSM in the greater Bay Area revealed the most dependable districts, areas where class omission was frequent, and lack of structural connectivity among mapped vegetation spots. The weighted schemes performed reliably, yielding mean accuracy of 0.8372 and mean F1 score of 0.8267, whereas the unweighted setting made both measures worse. Heterogeneous OSM-derived labels, therefore, should be subject to reliability weighting.

The results of comparison of the classifiers provided basis for making planning-oriented conclusions. Specifically, W-SVM offered the most effective strategy by reducing the number of false positives, producing the lowest false positive rate and achieving the highest specificity and utility value. This conclusion justified the use of calibrated probability gating and an additional uncertain decision state to prevent overestimated supply of UGS.

City diagnostics helped determine cities with high omission pressure and those requiring more reliable detection. The mean precision of 0.885 was higher than mean recall of 0.759, implying conservative estimation with potential problems in some cities. Comparison of land cover types provided the basis for the third conclusion – UGS area was not constant across products since the 10 m estimate of 139,427.06 ha lay between ALCC and CLCD. The MSPA statistics led to the fourth conclusion – core-dominated but connector-poor morphology of the green landscape.

Thus, the overall conclusion drawn from these analyses was that a reliable map of the greater Bay Area's green

spaces should be evaluated from four aspects – classification map, detectability surface, area estimate, and ecological structure. Neither accuracy nor area measure could help to identify cities with omission errors, detect disagreement in products, and assess ecological connectivity. The reliability-weighted conformal graph calibration helped resolve these issues and retained the basis of evidence – Sentinel-2 and OSM data.

References

[1] Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing, 114*, 24-31.

[2] Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing, 65*(1), 2-16.

[3] Bolund, P., & Hunhammar, S. (1999). Ecosystem services in urban areas. *Ecological Economics, 29*(2), 293-301.

[4] Bowler, D. E., Buyung-Ali, L., Knight, T. M., & Pullin, A. S. (2010). Urban greening to cool towns and cities: A systematic review of the empirical evidence. *Landscape and Urban Planning, 97*(3), 147-155.

[5] Boykov, Y., Veksler, O., & Zabih, R. (2001). Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 23*(11), 1222-1239.

[6] Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32.

[7] Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794).

[8] Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment, 37*(1), 35-46.

[9] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273-297.

[10] Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., ... & Bargellini, P. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment, 120*, 25-36.

[11] Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment, 80*(1), 185-201.

[12] Gao, B. C. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment, 58*(3), 257-266.

[13] Gill, S. E., Handley, J. F., Ennos, A. R., & Pauleit, S. (2007). Adapting cities for climate change: the role of the green infrastructure. *Built Environment, 33*(1), 115-133.

[14] Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal, 69*(4), 211-221.

[15] Haase, D., Larondelle, N., Andersson, E., Artmann, M., Borgström, S., Breuste, J., ... & Elmqvist, T. (2014). A quantitative review of urban ecosystem service assessments: concepts, models, and implementation. *Ambio, 43*(4), 413-433.

[16] Haklay, M., & Weber, P. (2008). Openstreetmap: User-generated street maps. *IEEE Pervasive Computing, 7*(4), 12-18.

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

[18] Heipke, C. (2010). Crowdsourcing geospatial data. *ISPRS Journal of Photogrammetry and Remote Sensing, 65*(6), 550-557.

[19] Herold, M., Goldstein, N. C., & Clarke, K. C. (2003). The spatiotemporal form of urban growth: measurement, analysis and modeling. *Remote Sensing of Environment, 86*(3), 286-302.

[20] Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment, 25*(3), 295-309.

[21] McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing, 17*(7), 1425-1432.

[22] McGarigal, K. (1995). FRAGSTATS: spatial pattern analysis program for quantifying landscape structure (Vol. 351). US Department of Agriculture, Forest Service, Pacific Northwest Research Station.

[23] Myers, V. I., Westin, F. C., Horton, M. L., & Lewis, J. K. (1974). Effective use of ERTS multisensor data in the Northern Great Plains (No. SDSU-RSI-74-09).

[24] Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., & Wulder, M. A. (2014). Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment, 148*, 42-57.

[25] Platt, J. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in Large Margin Classifiers, 10*(3), 61-74.

[26] Pontius Jr, R. G., & Millones, M. (2011). Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing, 32*(15), 4407-4429.

[27] Soille, P., & Vogt, P. (2009). Morphological segmentation of binary patterns. *Pattern Recognition Letters, 30*(4), 456-459.

[28] Stehman, S. V. (1997). Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment, 62*(1), 77-89.

[29] Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment, 8*(2), 127-150.

[30] Tzoulas, K., Korpela, K., Venn, S., Yli-Pelkonen, V., Kaźmierczak, A., Niemela, J., & James, P. (2007). Promoting ecosystem and human health in urban areas using Green Infrastructure: A literature review. *Landscape and Urban Planning, 81*(3), 167-178.

[31] Van den Bosch, M., & Sang, Å. O. (2017). Urban natural environments as nature-based solutions for improved public health–A systematic review of reviews. *Environmental Research, 158*, 373-384.

[32] Vogt, P., Riitters, K. H., Estreguil, C., Kozak, J., Wade, T. G., & Wickham, J. D. (2007). Mapping spatial patterns with morphological image processing. *Landscape Ecology, 22*(2), 171-177.

[33] Vovk, V., Gammerman, A., & Shafer, G. (2005). Algorithmic learning in a random world. Boston, MA: Springer US.

[34] Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing, 27*(14), 3025-3033.

[35] Yang, J., & Huang, X. (2021). The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019. *Earth System Science Data, 13*(8), 3907-3925.

[36] Yuan, B., Wan, Z., Wu, L., Zhang, A., Yang, X., Li, X., & Chen, C. (2026). Urban Green Space Mapping from Sentinel-2 and OpenStreetMap via Weighted-Sample SVM Classification. *Remote Sensing, 18*(2), 272.

[37] Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., ... & Arino, O. (2022). ESA WorldCover 10 m 2021 v200.